

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: January 13, 2004, 05:14:36 ; Search time 4187 Seconds
(without alignments)
11383.109 Million cell updates/sec

file: US-09-922-225A-1
effect score: 1961
sequence: 1 agccttgccatgctgctgc.....aggtctctgacatgaaaaa 1961

coring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

searched: 22781392 seqs, 12152238056 residues

total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_esti:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_estfun:*

15: em_estom:*

16: em_gss_hum:*

17: em_gss_inv:*

18: em_gss_pin:*

19: em_gss_vrt:*

20: em_gss_fun:*

21: em_gss_man:*

22: em_gss_mus:*

23: em_gss_pro:*

24: em_gss_rod:*

25: em_gss_phg:*

26: em_gss_vrl:*

27: gb_gss1:*

28: gb_gss2:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1444.4	73.7	2036	11 AK054293	AK054293 Mus muscu
2	1442.8	73.6	2586	11 AK050335	AK050335 Mus muscu
3	1435	73.2	2591	11 AK086674	AK086674 Mus muscu
4	977.4	49.8	1201	9 ALS16558	ALS16558 ALS16558

C	5	931.4	47.5	972	13	BX383922	BX383922
	6	879.6	44.9	1160	13	BX345346	BX345346
	7	872.2	44.5	1006	13	BX383923	BX383923
C	8	869.4	44.3	1201	9	ALS16557	ALS16557
	9	852.8	43.5	891	13	BX329238	BX329238
	10	850.4	43.4	1080	13	BX361838	BX361838
	11	816.4	41.6	933	13	BX369641	BX369641
C	12	813.6	41.5	942	9	ALS76475	ALS76475
	13	809.2	41.3	851	13	BU186063	BU186063
	14	802.8	40.9	1016	12	BM450315	BM450315
	15	791.6	40.4	856	13	BU132994	BU132994
	16	790.6	40.3	859	12	BI832816	BI832816
	17	790	40.3	1054	12	BI092420	BI092420
	18	787.8	39.6	873	14	CD358784	CD358784
	19	775.8	39.6	873	14	CD358784	CD358784
	20	774	39.5	1037	12	BM468168	BM468168
	21	772.4	39.4	853	14	CD300154	CD300154
C	22	749	38.2	771	13	BU739289	BU739289
	23	742	37.8	872	13	BU167220	BU167220
	24	730.4	37.2	824	12	BM456432	BM456432
	25	726	37.0	855	12	BI825443	BI825443
	26	722.4	36.8	946	10	EG674191	EG674191
	27	722.4	36.8	979	10	EG681522	EG681522
	28	722.2	36.8	1048	13	BQ955364	BQ955364
	29	717.6	36.6	1154	12	BM906750	BM906750
	30	713.4	36.4	960	10	BF795352	BF795352
	31	710	36.2	751	10	EG701766	EG701766
	32	700.6	35.7	1102	12	BM906325	BM906325
	33	687.2	35.0	924	12	BM451188	BM451188
	34	686.2	35.0	1120	12	BM805877	BM805877
	35	684.4	34.9	752	12	BI772096	BI772096
C	36	684	34.9	698	13	BU623771	BU623771
	37	681.6	34.8	775	10	EG717364	EG717364
	38	675.8	34.5	953	10	BF795135	BF795135
	39	673.2	34.3	939	10	BF795137	BF795137
	40	661.8	33.7	1182	12	BM545693	BM545693
	41	650.4	33.2	653	12	BM714659	BM714659
	42	648	33.0	713	10	EG717946	EG717946
	43	626.4	31.9	810	12	RG961872	RG961872
	44	626	31.9	682	14	CB155405	CB155405
	45	620.6	31.6	759	12	BM944500	BM944500

RESULT 1

AK054293

LOCUS

DEFINITION

Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN

full-length enriched library, clone:E33001F20 product:D1BD1

homolog [Homo sapiens], full insert sequence.

AK054293

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

AK054293 2036 bp mRNA linear HTC 05-DEC-2002
Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
full-length enriched library, clone:E33001F20 product:D1BD1
homolog [Homo sapiens], full insert sequence.

AK054293

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

ALIGNMENTS

AK054293 2036 bp mRNA linear HTC 05-DEC-2002
Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
full-length enriched library, clone:E33001F20 product:D1BD1
homolog [Homo sapiens], full insert sequence.

AK054293 GI:26344126
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

REFERENCE
AUTHORS

3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamanoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

20530913
11076861
4 Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Iehii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,Y., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaide,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Mashio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Bash,S., Blake,J., Brownstein,M.J., Bult,C., Carninci,P., de Boraio,M.F., Brownstein,M.J., Gustincich,S., Hill,D., Fletcher,C., Fujita,M., Gariboldi,M., Lee,N.H., Lyons,P., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Whitaker,C., Wilming,L., Wyshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,J., Kohsaki,S. and Hayashizaki,Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

21085660
11217551
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS

6 (bases 1 to 2036)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hangaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saichoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akaira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN) Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

Location/Qualifiers

SOURCE

1. 2036
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="E330011F20"
/db_xref="FANTOM_DB:E330011F20"
/db_xref="taxon:10090"
/clone="E330011F20"
/sex="female"
/tissue type="ovary"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
67. 1902
/note="unnamed protein product; DIBDI homolog [Homo sapiens] (SPTR|Q96PB9, evidence: FASTY, 93.2%ID, 100%length, match=1833)
putative"
/codon_start=1
/protein_id="BAC35720.1"
/db_xref="GI:26344127"
/translation="MASRRARQRKGGGGGGGAGAAEKLQSGSEAGAEPRP
ESNGKAGQWAPGEGTAPKCLLSARLCAALLNSDCDETFNWEPTHYLYYKGFQT
WEKSPVARSYAYLLHAPPAARILQTNKILVYFLRLLAFVSCVCELYKGFYKA
VCKEGLHVSRLMFLVLTSGMFCSSSAFLPSFCMYTTLIAMTYMDYVYVIG
VAAGAILGWPFSALGLPIADLLARKRWKSFLLSLVALFLVPMVVDYVYIG
LVAPNLIVNVFTSHGPDLYGTEPWYFLINGFLNFVAFALALPLTILMEYL
LORFVQNLGHPYWTLPAPYIMFIPIOPHKEERFLFPVPLICLCAVALSALOK
CTHFVQRIRLHYTIVSNWALGTVELGLLSRSVALPFGVHGLDLYPPEYRIA
TDPTHTTPGRRPVNVCVKWEYRFPSPFLPDNWQLQFIPSPERGQIPKPRAGPLA
TTTPTVTHNDQNRPEPSYIDISKHYLVLDITRETPRPNPYNSHREWWSLAHREF
LDASRSSKLLAFYVFLSDQYTVVYVNTILKPRKXKPSRKSGG"
BASE COUNT 427 a 542 c 528 g 539 t

ORIGIN

Query Match 73.7%; Score 1444.4; DB 11; Length 2036;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1690; Conservative 0; Mismatches 251; Indels 25; Gaps 7;

QY 1 AGGCTTGCCATGCTAGTCGAGGGGCTCGGAGCGCCTGAAGGGCAGCGGGGCCA----- 55
Db 58 ACCTGGGCGATGCGGAGCGGAGGCGGCGGAGCGCCTCAAGGGTGGCGGGCGGTGGC 117
QY 56 -----GCAGTGGGATACGCCCGCGCTGGCGACAGCTCGCGAGCTCTGGCGAGCGCA 111
Db 118 GGGCGGCGGTGGGATCGCGCGCGCGCGGAGAGAGCT---GGAACAGCTCGCGAGCGCG 174
QY 112 GAGCGCGCGCGCGCGGAGCACCGGACCGAGTATCTGGGAAACAAAGCAGGCAAGCTCTGG 171
Db 175 GAGGC---AGGCCGAGCGCGCGCGCGCGAGAG---TCTGGAAATAAAGCAGGCGAAGCTCTGG 228
QY 172 GCACCTGAAGGATCTACTGCTTTCAAGTGTCTGCTTTTCAGCAAGGTTATGCTGCTCTC 231
Db 229 GCCCCTGAAGGATCCACTGCTTTCAAGTGTCTGCTCTCAGCGAGGTTGTGCGCAGCCCTT 288
QY 232 CTGAGCAACATCTCTGACTGTGTATGAAACATCAACTACTCGGAGCGCAACACACTACCTC 291
Db 289 CTGAGCAACATCTCGGACTGTGTATGAAACATCAACTACTCGGAGCGCAACACACTACCTC 348
QY 292 ATCTATCGGGAAGGTTTCAGACTTCGGAAATATTCAGCAATATGCCATTCGCTCTTAT 351
Db 349 ATCTACGGAAGGATTCAGACATCGGAATATTCGCCGTGTATGCCATTCGCTCTTAT 408
QY 352 GCTTACCTGTGCTTCACTGCTCGCGAGCTGCAATTCATGCAAGGATTCACAACTAAT 411
Db 409 GCTTACCTGTGCTGCGACGCTGCGCGAGCTGCAATTCATGCAAGGATTCGCGAGCTTAC 468
QY 412 AAGATCTCTGTGTTTACTTTTTCGAGTGTCTTCTGGCTTTTGTGAGCTGTATTGTGAA 471
Db 469 AAGATCTCTGTATTCTACTTTTGTGAGATGCTCTCTGCTTTTGTGAGCTGTATTGTGAA 528
QY 472 CTTTACTTTTACAGGCTGTGTCGAAGAGTTGGGTTGCACCTGAGTGAATGATGCTA 531
Db 529 CTTTACTTTTATAAGGCTGTGTGCAAGAGTTTGGGCTGCAATGTGAGTGAATGATGCTG 588

2y	532	GCCTCTTGGTTCTCAGACCTGGCATGTTTGTCTCANTCAGCATTCCTTCTTAGTAGC	591
2b	589	GCCTTCTCTGGTGTCTCAGCAGCGGCATGTTCTGTCTCATCTCATCTGATTCCTCCCCAGCAGC	648
2y	592	TTTCTGATGTACACTACGTTTGATAGCCATGACTTGGATGATATGGAACAAGACTTCCATT	651
2b	649	TTTCTGCATGTACACACGCTGATCGCCATGACGGGATGGTATATGACACAGACACCCATT	708
2y	652	GCTGTGTGGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTCAGTGCAGTCTTTGGT	711
2b	709	GCTGTGTGGGAGTAGCAGCTGGTGCCATCTTAGGCTGGCCGTTTAGTGCAGCTCTTGGT	768
2y	712	TTACCCATTGCTTTGATTTGCTGTGTTCATGAAACACAGGTGGAAGAGTTTCTTTTCATTGG	771
2b	769	TTACCANTCGGCTTCGACTTGTCTGGCCAGGAGGACAGGTGGAAGAGTTTCTCTCTG	828
2y	772	TGCTGTATGGCCCTCATACATATTTCTGGTCTGTGGTGTCATTGACAGCTACTATTAT	831
2b	829	TCCCTGGTGGCCCTGGCTCTCTTCTGGTGGCCGTGTGTCAATGACAGCTACTATTAC	888
2y	832	GGGAGTTTGGTGATTGCACACTTCACANTGTTTGTATTAATGCTTTTACTCTCTCATGGA	891
2b	889	GGGAGTTTGGTGGTGGCCCTTCACATGTTTGTATAGCTCTTTACGCTCTCATGGA	948
2y	892	CGTGATCTTTATGTTACAGAAACCCCTGCTATTCTCTTAATTAATGGAATTTCTGAATTC	951
2b	949	CTGTATCTTTATGGTACAGAACCATGTTTCTTAATCAATGAGATTTCTGAACCTTC	1008
2y	952	AATGATGCTTTGCTTTGGCTCTCTAGTCTTACACTGACTCTCTTTATGGAATACCTG	1011
2b	1009	AATGTGTGCTTTGGCTTTGGCTCTCTGGTTTGGCCACTGACTCTCTTTATGGAGTACCTG	1068
2y	1012	CTGCAGAGATTTCAATGTCAGAAATTAGGCCACCGCTATTGGCTTACCTTGGCTCCAATG	1071
2b	1069	CTGCAGAGTTTCCATGTTCAGAAATTAGGACACCGTATTGGCTGACCTTGGCGCCATG	1128
2y	1072	TATATTTGGTTTATAATTTTCTTATCCAGCCCTCAAAAGAGGAGAGATTTCTTTCCCT	1131
2b	1129	TATATTTGGTTTATAATTTTCTTATCCAGCCCTCAAAAGAGAGCGCTTTCTCTTCCCA	1188
2y	1132	GTGTATCCACTATATGCTCTGTGGCGCTGGGCTCTCTCTGCACTTTCAGAAATGTAC	1191
2b	1189	GTCTATCCGCTCATATGCGCTCTGTGGTGTGTTGTCTTCTTGCACTTCAGAAATGTAC	1248
2y	1192	CACCTTGTGTTTCAACGATATCGCTGGAGCACTATACGTGACATCGAATTGGCTGGCA	1251
2b	1249	CACCTTGTGTTTCAACGATATCGCTTGGAGCAATTACCGGTGACGTCCAACTGGCTGGG	1308
2y	1252	TTAGGAACGTGCTTCCGTGTTTGGGCTCTTGTGATTTTCTTCGCTCTGTGGCACTGTTAGA	1311
2b	1309	CTGGGACAGTTTTCCTGTGTTTGGGCTCTGTGTGTTTCTCCCGTCTGTGGCTCTGTTAGA	1368
2y	1312	GGATATCACGGGCCCTTGATTGTATCCAGAAATTTTACCGAATTTGCTACAGACCCCAACC	1371
2b	1369	GGATACCAAGGGCCCTTGATTTTATATCCAGAAATTTACAGAAATTTGCCACAGACCCAA	1428
2y	1372	ATCCACACTGTGCCAGAGAGGACGACTGTGAATGTCTGTGGGAAAAGATGGTATCGA	1431
2b	1429	ATCCACACTGTGCCAGAGAGGAGCCCGTGAAGCTGTGTGGGAAAAGATGGTATCGA	1488
2y	1432	TTTCCGACGAGCTTCCCTTCTTCGTGCAATTTGGCAGCTTCAGTTTCATTCATCAGAGTTC	1491
2b	1489	TTTCCAGTAGTCTTCCCTTCTGCTTGACANTTTGGCAGCTTCAGTTTCATTCATTCAGAGTTC	1548
2y	1492	AGAGGTTCAGTTTACAAAACTTTTTCGAGAAGGACCTCTGGCCACCCGAGTTTGTCTTACT	1551
2b	1549	AGAGGTTCAGTTTACAAAACTTTTTCGAGAAGGACCACTGGCCACCAAGGACTGTCTCCACT	1608
2y	1552	GACATGAATGACCAAACTAGAGAGGCCATCCAGATATATTGATATCAGTAAATGCAAT	1611
2b	1609	CACATGAACGACCAAACTAGAGAGGAGCCCTTCAGATATCATTGACATCAGCAATGCAAT	1668
2y	1612	TATTTAGTGGATTTTGACACCATGAGAGAAACACCCCGGAGGCCAAAATATTTCATCCAAT	1671

Db	1669	TACTTTAGTGGGATTTGGATACCATGAGAGAAACACCTGGGAGCCAAACTACTCATCCAC	1728
QY	1672	AAAGAAGAATGGATTCAGCTTGGGCTATAGACCACTTCCTTTGATGCTTCTAGATCTTCAAAG	1731
Db	1729	ACAGAAGAAGTGGGTACGCTGGCTCACAGACCACTCTGGATCGGTCTAGGTCCTTCCAAG	1788
QY	1732	CTGCTGGGGGCAATCTATGTCCTCTCCCTGTCAGATCAGTATACAGTCTAGCTAACTAC	1791
Db	1789	TTGCTTCGGGCATTTCTATGTTCTCTCTCTCTGTCAGATCAGTACACGGTGTATGTGACTAC	1848
QY	1792	ACCATCTCTCAACCCCGGAAAGCAAAAGCAAAATCAGGAGAAAGTGGAGGTTAGCAACAC	1851
Db	1849	ACCATCTCTGAAGCTTCGGAAGCAAAAGCCGAGCAGGAAGAAAGTGGAGGTTAG---CAC	1905
QY	1852	ACCTGTGGCCCAAGAGCAACACATCTTGTTAACATTAATTGATTCAGTCACTGACTCCCT	1911
Db	1906	CCCGGTGACTCCAGAGACAGCC--TGTGCTAACTAGGGCTTCCACGGAGCGGACTTCCT	1963
QY	1912	GCAAGTCATCGCTGTAACTTTGTATATAAAGGTCCTTCGACATCA	1957
Db	1964	CCAAAGGCAAT--CTGTAGAGCTGTATAAAGTCTGTCCACCAAGA	2007
RESULT 2			
AK050335		2586 bp	linear
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
REFERENCE			
AUTHORS			


```
1132 GTGTATCCACTATATGTCTCTGTGGCGTGTGGCTCTCTGTGCACTTCAGAAATGTTAC 1191
1223 GTCTATCCGCTCATATGCTCTGTGGTGTGTGTCTCTCTGTGCACTTCAGAAATGTTAC 1282
1192 CATTGTGTGTTTCAAGATATCGCTGGAGCACTATCTGTGACATCGAATGCTGGCA 1251
1283 CATTGTGTGTTTCAAGATATCGCTGGAGCACTATCTGTGACATCGAATGCTGGCG 1342
1252 TTAGGAAGTCTCTCTCTGTGTTGGCTCTTGTCTATTTCTCTGCTGTGGCACTGTTTACA 1311
1343 CTGGGACAGCTTTCTCTGTGTTGGCTCTTGTCTCTCTGCTGTGGCTGTGTTTACA 1402
1312 GATATACACGGGCGCTTGTATTTGTATCCAGATTTTACGATTTTACGATTTTACGATTTT 1371
1403 GATATACACGGGCGCTTGTATTTGTATCCAGATTTTATCCAGATTTTACGATTTTACGATTTT 1462
1372 ATCCACACTCTCCAGAAAGCAGACCTGTGAATGTCTGTGGGAAAAGAGTGGTATCGA 1431
1463 ATCCACACTCTCCAGAAAGCAGACCTGTGAATGTCTGTGGGAAAAGAGTGGTATCGA 1522
1432 TTTCAGAGAGCTTCTCTCTCTGTGCAATTTGCGAGCTTCAGTTCATTTCCATCAGAGTTC 1491
1523 TTTCAGAGAGCTTCTCTCTCTGTGCAATTTGCGAGCTTCAGTTCATTTCCATCAGAGTTC 1582
1492 AGAGTCTAGTATACCAAAACCTTTTGCAGAGAGGACCTCTGGCCACCGGATTTGTTCTACT 1551
1583 AGAGTCTAGTATACCAAAACCTTTTGCAGAGAGGACCTCTGGCCACCGGATTTGTTCTACT 1642
1552 GATGATGACAGCAATCTAGAGAGGACCTTCAGATATTTGATATCGATTAATGCTAT 1611
1643 CACATGAACAGCAGAACCCGAGAGGAGCTTCAGATATTTGATATCGATTAATGCTAT 1702
1612 TATTTAGTGTGATTTGGACACCATGAGAGAACCC - CGGAGGACCAAAATATTCATCAA 1670
1703 TACTTGTGATTTGGATACATGAGAGAACCCCTCGGAGGCCAACTACTCATCCCA 1762
1671 TAAAGAGATGATGATGATGCTTGTGCTATGACATTTGATGCTTGTGATGCTTGTGAT 1730
1763 CAGAGAGAGTGGGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1822
1731 GCTCTGCGGCACTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1790
1823 GTTCTGCGGCACTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1882
1791 CACATCTTCAAAACCCGAGAGCAAAAGCAAAATCAGAGAAAGAGTGGAGTTAGCAACA 1850
1883 CACATCTTCAAAACCCGAGAGCAAAAGCAAAATCAGAGAAAGAGTGGAGTTAG ---CA 1939
1851 CACCTGTGGCCCAAGAGCAAAATCTGTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1910
1940 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1977
1911 TGCAAGTCACTGCTGTAACATTTGTAATTAAGGCTCTTCTGACATCA 1957
1998 TCCAAGGCT--CTGTGTAGAGCTGTAATAAAGGCTGCTCCACCAAGA 2042
```

RESULT 4

AL516558

LOCUS

DEFINITION

AL516558 Homo sapiens

CSODA006YN06 5-PRIME, mRNA sequence.

ACCESSION

AL516558

VERSION

AL516558.2

GI:30491310

EST.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1201)

Li W.B., Gruber C., Jesses, J. and Polayes, D.

Full-length cDNA libraries and normalization

JOURNAL

Unpublished

COMMENT

On Feb 13, 2001 this sequence version replaced gi:12780051.
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODA006D03QPI&cluster=5955.r. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODA006D03QPI.

FEATURES

source

1. .1201

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODA006YN06"

/tissue_type="NEUROBLASTOMA"

/clone_lib="Homo sapiens NEUROBLASTOMA"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched into

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

Library was not normalized."

BASE COUNT

247 a

257 c

274 g

363 t

60 others

ORIGIN

Query Match

Best Local Similarity

Matches 1027; Conservative

49.8%; Score 977.4; DB 9; Length 1201;

94.1%; Pred. No. 2.1e-225;

35; Mismatches 22; Indels 7; Gaps 5;

10 ATGCTAGTTCGAGGGGCTCGGAGGCTCTGAAGGGCAGCGGGCCAGCAGTGGGATACG 69

50 ATGCTAGTTCGAGGGGCTCGGAGGCTCTGAAGGGCAGCGGGCCAGCAGTGGGATACG 109

70 GCGGCGGCTGCGGAGCAAGCTGCGGGAGCTGCTGGGAGCAGCGGAGCGGGCGGAG 129

110 GCGGCGGCTGCGGAGCAAGCTGCGGGAGCTGCTGGGAGCAGCGGAGCGGGCGGAG 169

130 CACGAGCAGGAGTTCCTGGGAGCAAGCAGGAGCAAGTCTGGGAGCCTGAAGGATCTACT 189

170 CACGAGCAGGAGTTCCTGGGAGCAAGCAGGAGCAAGTCTGGGAGCCTGAAGGATCTACT 229

190 GCTTCAAGTGTCTGCTTTTCAAGAGTTATGCTGCTCTCTGAGCAACATCTCTGAC 249

230 GCTTCAAGTGTCTGCTTTTCAAGAGTTATGCTGCTCTCTGAGCAACATCTCTGAC 289

250 TGTGATGAACATTCACATCTGAGGAGCAACACACTACCTCATCTATGGGAGAGGTTT 309

290 TGTGATGAACATTCACATCTGAGGAGCAACACACTACCTCATCTATGGGAGAGGTTT 349

310 CAGACTTGGGAATATTCCCGCAGCATATGCCATTGCTCTCTATGTTTACCTGTTTCAAT 369

350 CAGACTTGGGAATATTCCCGCAGCATATGCCATTGCTCTCTATGTTTACCTGTTTCAAT 409

370 GCGTGGCAGCTGCATTTTCATGCAAGATTTTACAACTTAATAGATCTTGTGTTTAC 429

410 GCGTGGCAGCTGCATTTTCATGCAAGATTTTACAACTTAATAGATCTTGTGTTTAC 469

430 TTTTTCGATGCTTCTGCTGCTTTTGTGAGCTGCTATTGTGAACTTTTACTTTTACAAGCT 489

470 TTTTTCGATGCTTCTGCTGCTTTTGTGAGCTGCTATTGTGAACTTTTACTTTTACAAGCT 529

490 GTGTGCAAGAGTTTGGTGTGACGTGAGTCAAAATGATGCTAGCTTCTTGTGTTTAC 549

530 GTGTGCAAGAGTTTGGTGTGACGTGAGTCAAAATGATGCTAGCTTCTTGTGTTTAC 589

550 ACTGGCATGTTTTGCTCATCATCAGCAATTCCTTCTAGTAGCTTCTGTATGTACTACG 609

590 ACTGGCATGTTTTGCTCATCATCAGCAATTCCTTCTAGTAGCTTCTGTATGTACTACG 649

610 TTGATAGCCATGACTGGATGATATATGACCAAGACTTCCATTGCTGCTGGAGTAGCA 669
650 TTGATAGCCATGACTGGATGATATATGACCAAGACTTCCATTGCTGCTGGAGTAGCA 709
670 GTGGGGCTACTTAGGCTGGCCATTCAGTGCAGCTCTTGTTTACCCATTGCTTTGAT 729
670 GTGGGGCTACTTAGGCTGGCCATTCAGTGCAGCTCTTGTTTACCCATTGCTTTGAT 769
730 TTGCTGGTCAATGAACACAGGTGGAAGATTTCTTTCAATGGTGGTGTATGCGCCCTCAT 789
770 TTGCTGGTCAATGAACACAGGTGGAAGATTTCTTTCAATGGTGGTGTATGCGCCCTCAT 829
790 CTATTTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 849
830 CTATTTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 889
850 CCACCTCAACATTTGTTGATATATGCTCTTTACTCCTCATGAGCCTGATCTTTATGTTACA 909
890 CCACCTCAACATTTGTTGATATATGCTCTTTACTCCTCATGAGCCTGATCTTTATGTTACA 949
910 GAACCTGGTATTTCTATTTAATTAATGATTTCTGAATTCATGAGCCTTTGCTTTG 969
950 GAACCTGGTATTTCTATTTAATTAATGATTTCTGAATTCATGAGCCTTTGCTTTG 1009
970 GCTCTCTAGTCTACCTGACTCTCTCTTTATGGAATACCTGCTGCAGAGATTTGATTT 1029
1010 GCTCTCTAGTCTACCTGACTCTCTTTATGGAATACCTGCTGCAGAGATTTGATTT 1066
1030 CAGATTTAGGCGACCGTATGCTTACCTGCTGCAGAGATTTGATTTGATTTAATTT 1089
1067 CAGATTTAGGCGACCGTATGCTTACCTGCTGCAGAGATTTGATTTGATTTAATTT 1122
1090 TTCTCTCATCCA 1100
1123 TVAHCMCCM 1133

RESULT 5
X383922/c
CCUS
EFINITION
CESSION
ERSON
EWORDS
OURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK002BG03NP1&cluster=5955.r. Contact :
Peng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DK002BG03NP1.
Location/Qualifiers
1. .972
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK002YMO6"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"

/clone lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and BclI
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 269 a 196 c 242 g 247 t 18 others
ORIGIN
Query Match 47.5%; Score 931.4; DB 13; Length 972;
Best Local Similarity 98.3%; Pred. No. 2.7e-214;
Matches 943; Conservative 6; Mismatches 9; Indels 1; Gaps 1;
974 TCCTAGTCTTACCACTGACTCTCTTATGGAATACCTGCTGCAGAGATTTCAATGTTTCA 1033
972 TCCTAGTCTTACCACTGACTCTCTTATGGAATACCTGCTGCAGAGATTTCAATGTTTCA 913
1034 ATTTAGGCCACCGTATGCTTACCTTGGCTCCCAATGATATTTGGTTTATAAATTTCT 1093
912 ATTTAGGCCACCGTATGCTTACCTTGGCTCCCAATGATATTTGGTTTATAAATTTCT 853
1094 TCATCCAGCCTCACAAGAGAGAGAGATTTCTTTCCCTGCTGATCCACTTATATGTTCTCT 1153
852 TCATCCAGCCTCACAAGAGAGAGAGATTTCTTTCCCTGCTGATCCACTTATATGTTCTCT 793
1154 GTGGCGTGTGCTCTCTCTGCACTTCAGAAATGTTACCACTTTGTTTTCACAGATATC 1213
792 GTGGCGTGTGCTCTCTCTGCACTTCAGAAATGTTACCACTTTGTTTTCACAGATATC 733
1214 GCCTGGAGCACTATACCTGACATCGAATTTGGCTGGCAATAGGAATGCTCTTCCTGTTTG 1273
732 GCCTGGAGCACTATACCTGACATCGAATTTGGCTGGCAATAGGAATGCTCTTCCTGTTTG 673
1274 GGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTCAGAGATATCAGGGC-CCCTTGAT 1332
672 GGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTCAGAGATATCAGGGCACCCTTGAT 613
1333 TTGTATCCAGAAATTTTACCGAATTTGCTACAGACCCCAACATCCACACTGTCACAGAGGC 1392
612 TTGTATCCAGAAATTTTACCGAATTTGCTACAGACCCCAACATCCACACTGTCACAGAGGC 553
1393 AGACTGTGAATGCTCTGTGGGAAAGAGTGTATCGAATTTCCAGAGAGCTTCTCTTCTT 1452
552 AGACTGTGAATGCTCTGTGGGAAAGAGTGTATCGAATTTCCAGAGAGCTTCTCTTCTT 493
1453 CCTGACAAATGGCAGCTTCACTGATTCATTCATCAGAGTTCAGAGCTCAGTTTACCAAACT 1512
492 CCTGACAAATGGCAGCTTCACTGATTCATTCATCAGAGTTCAGAGCTCAGTTTACCAAACT 433
1513 TTTGAGAGAGGACCTCTGGCCACCCGATTTGTTCTTACTGACATGAATGACCAAGATCTA 1572
432 TTTGAGAGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 373
1573 GAAGAGCCTCCAGATATATGATATCAGTAAATGCAATTTAGTGGATTTGGACACC 1632
372 GAAGAGCCTCCAGATATATGATATCAGTAAATGCAATTTAGTGGATTTGGACACC 313
1633 ATGAGAGAAACACCCCGGAGCCAAAATATTTCATCCAAATGAAGAATGATGATGATG 1692
312 ATGAGAGAAACACCCCGGAGCCAAAATATTTCATCCAAATGAAGAATGATGATGATG 253
1693 GCCTATAGACCAATTCCTTGTATGCTTCTAGATCTTCAAGCTGCTGGGCACTTCTATGTC 1752
252 GCCTATAGACCAATTCCTTGTATGCTTCTAGATCTTCAAGCTGCTGGGCACTTCTATGTC 193
1753 CCCTTCTCTGTCAGATCAGTATACAGTGTAGTAACTTACACCATCTCTCAAAACCCCGGAA 1812
192 CCCTTCTCTGTCAGATCAGTATACAGTGTAGTAACTTACACCATCTCTCAAAACCCCGGAA 133
1813 GCAAGAGCAATCAGAGAGAAAGTGGAGGTAGCAACACACCTGTGCGCCCAAGAGCAA 1872
132 GCAAGAGCAATCAGAGAGAAAGTGGAGGTAGCAACACACCTGTGCGCCCAAGAGCAA 73
1873 CCATCTTGTAACTATTGATTTCCAGTGCCTGACTCCCTGCAAGTCAATCGCTGTAA 1931


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODK002YM06"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/clone_lib="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."
ASE COUNT 212 a 229 c 265 g 296 t 4 others
RIGIN

Query Match 44.5%; Score 872.2; DB 13; Length 1006;
Best Local Similarity 59.4%; Pred. No. 5.7e-200;
Matches 893; Conservative 2; Mismatches 1; Indels 2; Gaps 2;

Y 1 AGGCTTGCCATGGCTAGTCGAGGGCTCGGAGGCTGGAAGGCGCAGCGGGCGGCAGT 60
b 111 AGGCTTGCCATGGCTAGTCGAGGGCTCGGAGGCTGGAAGGCGCAGCGGGCGGCAGT 170
Y 61 GGGGATACGGCCCGGCTCGGACAAAGCTGGGAGCTGCTGGGACGCGAGGCGGC 120
b 171 GGGGATACGGCCCGGCTCGGACAAAGCTGGGAGCTGCTGGGACGCGAGGCGGC 230
Y 121 GGGGACGACGACGACGACGAGTATCTGGGACAAAGCAGGACGAGTCTGGGACCTGAA 180
b 231 GGGGACGGA-SAGCGGACGAGTATCTGGGACAAAGCAGGACGAGTCTGGGACCTGAA 269
Y 181 GGATCTACTGCTTTCAAGTGTCTGCTTTGAGCAAGGTTATGCTGCTCTCTCTGAGCAAC 240
b 290 GGATCTACTGCTTTCAAGTGTCTGCTTTGAGCAAGGTTATGCTGCTCTCTCTGAGCAAC 349
Y 241 ATCTCTGACTGTGATGAACATTTCAACTACTGGGACCAACACACTACTCATCTATGGG 300
b 350 ATCTCTGACTGTGATGAACATTTCAACTACTGGGACCAACACACTACTCATCTATGGG 409
Y 301 GAAGGTTTCAGACTTGGGAAATATCCCGACATATGCCATGCTGCTCTCTATGCTTACTG 360
b 410 GAAGGTTTCAGACTTGGGAAATATCCCGACATATGCCATGCTGCTCTCTATGCTTACTG 469
Y 361 TTGCTTCATGCTGCGGACGCTGCTTTCATGCAAGATTTCTACAACATAAAGATCTT 420
b 470 TTGCTTCATGCTGCGGACGCTGCTTTCATGCAAGATTTCTACAACATAAAGATCTT 529
Y 421 GTGTTTTACTTTTTGCGATGCTCTCTGCTTTGAGCTGATTTGGAATCTTACTTT 480
b 530 GTGTTTTACTTTTTGCGATGCTCTCTGCTTTGAGCTGATTTGGAATCTTACTTT 589
Y 481 TACAGGCTGTGTGCAAGAGTTTGGTTGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGT 540
b 590 TACAGGCTGTGTGCAAGAGTTTGGTTGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGT 649
Y 541 GTTCTCAGCACTGGCATGTTTGGCTCATCATCAGCAATTCCTTCTCTAGTCTCTGTATG 600
b 650 GTTCTCAGCACTGGCATGTTTGGCTCATCATCAGCAATTCCTTCTCTAGTCTCTGTATG 709
Y 601 TACACTACGTTGATGCCATGACTGATGGTATATGGACAAAGCTTCCATGCTGTGCTG 660
b 710 TACACTACGTTGATGCCATGACTGATGGTATATGGACAAAGCTTCCATGCTGTGCTG 769
Y 661 GGAGTAGCAGTGGGCTATCTTAGGCTGGCCATTCAGTGCAGCTCTTGGTTTACCCATT 720
b 770 GGAGTAGCAGTGGGCTATCTTAGGCTGGCCATTCAGTGCAGCTCTTGGTTTACCCATT 829
Y 721 GCCTTTGATTTGCTGTCATGAACACAGGTGAAGAGTTCTTTCATGTTGCTGCTGATG 780
b 830 GCCTTTGATTTGCTGTCATGAACACAGGTGAAGAGTTCTTTCATGTTGCTGCTGATG 889
Y 781 GCCTCTCATCTATTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
b 890 GCCTCTCATCTATTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 949

```

```

841 GTGATTGCACCACTCAACATGTTTGTATATGTTTACTCTCTCATGGACCTGATC 898
950 GTGATTGCA-CACTCAACATGTTTGTATATGTTTACTCTCTCATGGACCTGAT 1006

RESULT 8
AL516557/c 1201 bp mRNA linear EST 09-MAY-2003
LOCUS AL516557 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
DEFINITION CSODA006YN06 3-PRIME, mRNA sequence.
ACCESSION AL516557
VERSION AL516557.2 GI:30491308
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12780050.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODA006YN06&cluster=5955.r. Contact :
Feng Liang Email : fliang@lifetech.com URL : invitrogen.com/1600
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODA006DG03NP1.
FEATURES
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODA006YN06"
/tissue_type="NEUROBLASTOMA"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched.
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 354 a 217 c 270 g 297 t 63 others
ORIGIN

Query Match 44.3%; Score 869.4; DB 9; Length 1201;
Best Local Similarity 89.9%; Pred. No. 2.9e-199;
Matches 1009; Conservative 26; Mismatches 50; Indels 37; Gaps 9;

Y 816 TGACAGCTACTATTATGGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 875
b 1116 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1061
Y 875 CTTTACTCTCTCATGACCTGATCTTTATGTTGACAGAACCTTGGTATTTCTATTAATAA 935
b 1060 ATTTTACCTCTCATGACCTTCTTCTTATGACAGACCTTGGTATTTCTATTAATAA 1006
Y 936 TGGATTTCTGAATTTCAATGATGAGCTTTGCTTTGGCTCTCTCTCTCTCTCTCTCTCT 995
b 1005 --GGATTTCGATTTTAAATGATGAGCTTTT-STTTGGCTCTCTCTA-KCCTACACAT 951
Y 996 TCTTATGGATACCTGCTGCAGAGATTTTCATGTTTACAGATTTAGCCACCTGATTTGGCT 1055
b 950 TCTATGGATACCTGCTGCAGAGATTTTCATGTTTACAGATTTAGCCACCTGATTTGGCT 891
Y 1056 TACCTTGGCTTCCCAATGATATTTGGTTTATATATTTCTTATCCAGGCTCTCAAGAGGA 1115
b 1115 TACCTTGGCTTCCCAATGATATTTGGTTTATATATTTCTTATCCAGGCTCTCAAGAGGA

```


890 TACCTTGGCTCCAAATGATATATTTGGTTTATAATTTCTTTCATCCAGCCTCACAAAGAGGA 831
1116 GAGATTTCTTTTCCCTGTGTATCCACTTATATGTCTCTGTGGGCTGTGCTCTCTCTGC 1175
830 GAGATTTCTTTTCCCTGTGTATCCACTTATATGTCTCTGTGGGCTGTGCTCTCTCTGC 771
1176 ACTT-----CAGAAATGTTTACCACCTTTTGTG-TTTCACAGCATATC 1213
770 ACTTCAGMACAGTTTCTGTACTTCCAGAAATGTTACCACCTTTGTGTTTTCACAGCATATC 711
1214 GCCTGAGCAGCATATCTGTGATCATGAATTTGGCTGGGATTTAGGAACCTGCTTCTGTGTTG 1273
710 GCCTGAGCAGCATATCTGTGATCATGAATTTGGCTGGGATTTAGGAACCTGCTTCTGTGTTG 651
1274 GCCTTGTGCAATTTCTCGCTCTGTGGCACTGTTTCAGAGGATATCACGGGCCCTTGATT 1333
650 GCCTCTGTGCAATTTCTCGCTCTGTGGCACTGTTTCAGAGGATATCACGGGCCCTTGATT 591
1334 TGTATCCAGAAATTTTACCGAATTTGCTACAGACCCAAACCATCCACTGTCTCCAGAGGCA 1393
590 TGTATCCAGAAATTTTACCGAATTTGCTACAGACCCAAACCATCCACTGTCTCCAGAGGCA 531
1394 GACCTGTGAATCTCTGTGTGGGAAAGAGTGTATCGAATTTCCAGCAGCTTCTCTTCTTC 1453
530 GACCTGTGAATCTCTGTGTGGGAAAGAGTGTATCGAATTTCCAGCAGCTTCTCTTCTTC 471
1454 CTGACAAATGGCAGCTTCAGTTCATTCATCAGAGTTTCAGAGGTCAGTTACCAAAACCTT 1513
470 CTGACAAATGGCAGCTTCAGTTCATTCATCAGAGTTTCAGAGGTCAGTTACCAAAACCTT 411
1514 TTGCGAAGGAGCTCTGCGCCACCGGATTTCTTACTGATCAATGAATGACCAAGATCTAG 1573
410 TTGCGAAGGAGCTCTGCGCCACCGGATTTCTTACTGATCAATGAATGACCAAGATCTAG 351
1574 AAGAGCCATCCAGATATATTGATATCAGTAAATGCCATTTATTTAGTGGATTTGGACACCA 1633
350 AAGAGCCATCCAGATATATTGATATCAGTAAATGCCATTTATTTAGTGGATTTGGACACCA 291
1634 TGAGGAAACACCCCGGAGCCAAATTTTCATCCATAAGAGATGATCAGCTTGG 1693
290 TGAGGAAACACCCCGGAGCCAAATTTTCATCCATAAGAGATGATCAGCTTGG 231
1694 CCTATAGACCATTCCTTGATGCTTCTAGATCTTCAAAGCTGCTGCGGGCATTCTATGTCC 1753
230 CCTATAGACCATTCCTTGATGCTTCTAGATCTTCAAAGCTGCTGCGGGCATTCTATGTCC 171
1754 CTTTCTGTGATCAGTATACAGTGTACGTAACTACACCTCTCAAACCCCGGAAAG 1813
170 CTTTCTGTGATCAGTATACAGTGTACGTAACTACACCTCTCAAACCCCGGAAAG 111
1814 CAAAGCAATCAGGAGAAAGTGGAGGTTAGCAACACACCTGTGGCCCCCAAAGGACAAAC 1873
110 CAAAGCAATCAGGAGAAAGTGGAGGTTAGCAACACACCTGTGGCCCCCAAAGGACAAAC 51
1874 CATCTGTGTAAGTAACTA-TTGATTTCCAGTGTGATGCTGCTGCA 1914
50 CATCTGTGTAAGTAACTAANTANTNCCAGTGACCTGACTSCBNVAA 9

RESULT 9
LOCUS BX329238/c
DEFINITION BX329238 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
CDNA clone CS0DK002YM06 3-PRIME, mRNA sequence.
ACCESSION BX329238
VERSION BX329238.1 GI:30340822
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAI036ZC03_CS03379_1&cluster=5955.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAI036ZC03_CS03379_1.
Location/Qualifiers
1. 891
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="CS0DK002YM06"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with NotI and cloned into the NotI and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 314 a 193 c 190 g 193 t 1 others
ORIGIN
Query Match 43.58; Score 852.8; DB 13; Length 891;
Best Local Similarity 99.3%; Pred. No. 2.7e-195;
Matches 887; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 233 TGAGCAACATCTCTGACT-GTGATGAACATTTCACTACTGGGAGCCAAACACTACCTC 291
Db 891 TGAGCAACATCTCTGACTGGTGTGAACATTTCACTACTGGGAGCCAAACACTACCTC 832
QY 292 ATCTATGGGAGAGGTTTCAGACTTGGGATATTTCCCGACATATGCCATTCGCTCCAT 351
Db 831 ATCTATGGGAGAGGTTTCAGACTTGGGATATTTCCCGACATATGCCATTCGCTCCAT 772
QY 352 CTTTACCTGTTGCTTCACTGCTGCGCAGCTGCATTTTCATGCAAGAAATTTCAAACTAT 411
Db 771 CTTTACCTG-TGCTTCATGCTGCGCAGCTGCATTTTCATGCAAGAAATTTCAAACTAT 713
QY 412 AAGATTTCTGTGTTTACTTTTTCGCAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 471
Db 712 AAGATTTCTGTGTTTACTTTTTCGCAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 653
QY 472 CTTTACCTTTTACAAGGCTGTGTGCAAGAAATTTGGGTTGCTGCTGCTGCTGCTGCTGCT 531
Db 652 CTTTACCTTTTACAAGGCTGTGTGCAAGAAATTTGGGTTGCTGCTGCTGCTGCTGCTGCT 593
QY 592 GCCTTCTGTGTTCTCAGCACTGGCATGTTTGTGCTCATCATCAGCATTCCTTCTAGTAGC 591
Db 592 GCCTTCTGTGTTCTCAGCACTGGCATGTTTGTGCTCATCATCAGCATTCCTTCTAGTAGC 533
QY 592 TTCTGTATGTACACTAGCTGTATAGCCATGACCTGGAGTGTATATGCAAGAACTTCCATT 651
Db 532 TTCTGTATGTACACTAGCTGTATAGCCATGACCTGGAGTGTATATGCAAGAACTTCCATT 473
QY 652 GCTGTGTGGAGTAGCAGCTGGGCTATCTTAGGCTGGCCATTTAGTGCAGCTCTTGTGT 711
Db 472 GCTGTGTGGAGTAGCAGCTGGGCTATCTTAGGCTGGCCATTTAGTGCAGCTCTTGTGT 413
QY 712 TTACCCATTGCTTTGATTTGCTGCTCATGAAACACAGGTGGAGAGTTTCTTTCATTGG 771
Db 412 TTACCCATTGCTTTGATTTGCTGCTCATGAAACACAGGTGGAGAGTTTCTTTCATTGG 353
QY 772 TCCTGTATGGCCCTCATATCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 831
Db 352 TCCTGTATGGCCCTCATATCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 293

832 GGGAGTTGGTGAATGACCACTCAACATGTTGTTTGTATATGCTTTACTCTCATGGA 891
b 292 GGGAGTTGGTGAATGACCACTCAACATGTTGTTTGTATATGCTTTACTCTCATGGA 233
892 CTGTATCTTTATGTTACAGAACCCCTGGTATTTCTATTTAAATAATGGATTCTGAAATTC 951
b 232 CTGTATCTTTATGTTACAGAACCCCTGGTATTTCTATTTAAATAATGGATTCTGAAATTC 173
952 AATGTAGCCCTTTGCTTTGGCTCTCTAGTCTCTACACAGCTCTCTCTATGGAATACCTG 1011
b 172 AATGTAGCCCTTTGCTTTGGCTCTCTAGTCTCTACACAGCTCTCTCTATGGAATACCTG 113
1012 CTGACAGATTTTCATGTTTCAGAAATTTAGCCACCCGTTATTTGGCTTACCTTGGCTCCCAATG 1071
b 112 CTGACAGATTTTCATGTTTCAGAAATTTAGCCACCCGTTATTTGGCTTACCTTGGCTCCCAATG 53
1072 TATATTTGGTTTATATTTCTTCTATCAGCTCTCAAGAGGAGAGATTCT 1124
b 52 TATATTTGGTTTAT-ATTTTCTTCTATCAGCTCTCAAGAGGAGAGATTCT 1

RESULT 10
X361838 1080 bp mRNA linear EST 05-MAY-2003
EFINITION EX361838 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
CESSION Homo sapiens cDNA clone CS0DJ005Y120 5-PRIME, mRNA sequence.
ERSON BX361838
EYWORDS BX361838.1 GI:30380575
ORCE EST.
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1080)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5955.r for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DJ005BE10QPI&cluster=5955.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paraday Avenue Genoscope sequence ID : CS0DJ005BE10QPI.
Location/Qualifiers
1. 1080
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ005Y120"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
219 a 238 c 278 g 302 t 43 others

BASE COUNT 219 a 238 c 278 g 302 t 43 others
ORIGIN
Query Match 43.4%; Score 850.4; DB 13; Length 1080;
Best Local Similarity 98.2%; Pred. No. 1.1e-194; Indels 4; Gaps 4;
Matches 892; Conservative 8; Mismatches 4;
2 GGCTTGCCATGGCTAGTCGAGGGGCTCGGACGCGCTGAAGGGCGGCGGACAGCTG 61
|||||

Db 46 GGGATGCCATGGCTAGTCGAGGGGCTCGGACGCGCCTGAA - GGSAGCGGGCGCAGCAGTG 104
Qy 62 GGGATACGGCCCCGGCTGCGGACAAGCTGCGGAGCTGCTGGGACGCGGAGGGCG - GC 120
Db 105 GGGATACGGCCCCGGCTGCGGACAAGCTGCGGAGCTGCTGGGACGCGGAGGGCGGCG 164
Qy 121 GCGCGGAGCACCGGACCCGAGTTATCTGGGAAACAAAGCAGGACAAGTCTTGGGACCTGAA 180
Db 165 GCGCGGAGCACCGGACCCGAGTTATCTGGGAAACAAAGCAGGACAAGTCTTGGGACCTGAA 224
Qy 181 GGATCTACTGCTTTCAAGTCTGCTTTCAAGAGGTATGCTGCTCTCTCTCTGACCAAC 240
Db 225 GGATCTACTGCTTTCAAGTCTGCTTTCAAGAGGTATGCTGCTCTCTCTCTGACCAAC 284
Qy 241 ATCTCTGACTGCTGATGAAACATTCACTACTGGGAGCAACACACTACCTCATCTATGGG 300
Db 285 ATCTCTGACTGCTGATGAAACATTCACTACTGGGAGCAACACACTACCTCATCTATGGG 344
Qy 301 GAAGGTTTCAGCTTGGGAATATCCCGAGCATATGCCATTCGCTCCTATGCTTACCTG 360
Db 345 GAAGGTTTCAGCTTGGGAATATCCCGAGCATATGCCATTCGCTCCTATGCTTACCTG 404
Qy 361 TTGCTTTCATGCTGCGCCAGCTGCAATTTTCATGCAAGAAATTTACAAAATAAAGATTCTT 420
Db 405 TTGCTTTCATGCTGCGCCAGCTGCAATTTTCATGCAAGAAATTTACAAAATAAAGATTCTT 464
Qy 421 GTGTTTTCATGCTGCGCCAGCTGCAATTTTCATGCAAGAAATTTACAAAATAAAGATTCTT 480
Db 465 GTGTTTTCATGCTGCGCCAGCTGCAATTTTCATGCAAGAAATTTACAAAATAAAGATTCTT 524
Qy 481 TACAGGCTGTGTCAAGAAAGTTTGGGTTGCACGTGAGTCAAGTCAAGTCAAGTCAAGT 540
Db 525 TACAGGCTGTGTCAAGAAAGTTTGGGTTGCACGTGAGTCAAGTCAAGTCAAGTCAAGT 584
Qy 541 GTTCTCAGCACTGGGATGTTTGGTCTATCATCAGCAATTCCTTCTAGTAGCTTCTGTATG 600
Db 585 GTTCTCAGCACTGGGATGTTTGGTCTATCATCAGCAATTCCTTCTAGTAGCTTCTGTATG 644
Qy 601 TACACTACGTTGATAGCCATGACTGGAATGTTATGGACAAAGACTTCCATTGCTGTGCTG 660
Db 645 TACACTACGTTGATAGCCATGACTGGAATGTTATGGACAAAGACTTCCATTGCTGTGCTG 704
Qy 661 GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTAGTCAGCTCTTGTGTACCCATT 720
Db 705 GGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTAGTCAGCTCTTGTGTACCCATT 764
Qy 721 GCCTTTGATTTGCTGGTCAATGAAACACAGGTGGAGAGTTTCTTTTCAATTCGCTGATG 780
Db 765 GCCTTTGATTTGCTGGTCAATGAAACACAGGTGGAGAGTTTCTTTTCAATTCGCTGATG 824
Qy 781 GCCCTCATACATTTCT-GGTGCCTGTGGTGGTCAATGACAGCTACTATTATGGGAGTT 839
Db 825 GCCCTCATACATTTCTGGGTGCCTGTGGGCTCAATGACAGCTACTATTATGGGAGTT 884
Qy 840 GGTGATTCACCACTCAACATTTGTTTGTATATGCTTTTACTCTCT- CATGGACCTGATC 898
Db 885 GGTGATTCACCACTCAACATTTGTTTGTATATGKCTTTTACTCTCTTCAATGGGCTGATY 944
Qy 899 TTTATGGT 906
Db 945 TTTATGKT 952

RESULT 11
EX369641
LOCUS
DEFINITION EX369641 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK002YM06 5-PRIME, mRNA sequence.
ACCESSION EX369641
VERSION EX369641.1 GI:30449834
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Db 541 TCCCTGTGTATCCACTTATATGCTCTGTGTGGCGCTGTGGCTCTCTGTGCACCTTCAGAAAT 600
2Y 1187 GTTACCACTTTGTGTTTCAACAGATATCGCTCGAGCACTATCTGTGACATCGAATTGSC 1246
Db 601 GTTACCACTTTGTGTTTCAACAGATATCGCTCGAGCACTATCTGTGACATCGAATTGSC 660
2Y 1247 TGGCAATAGGAACCTGTCTTCCTGTTTGGGCTCTGTGCAATTTCTCGCTCTGTGGCACTGT 1306
Db 661 TGGCAATAGGAACCTGTCTTCCTGTTTGGGCTCTGTGCAATTTCTCGCTCTGTGGCACTGT 720
2Y 1307 TCAGAGATATCA-CGGGCCCTTGATTTGTATCCAGAAATTTACGAAATGCTACAGAC 1365
Db 721 TCAGAGATATCANCGGGCCCTTGATTTGTATCCAGAAATTTACGAAATGCTACAGAC 780
2Y 1366 CCAACCATCCACACTGTCCAGAGGAGCACTGTGAA---TGCTGTGTGGGAAAGAG 1422
Db 781 CCAACCATCCACACTGTCCAGAGGAGCACTGTGNAATGCTGTGTGGGAAAGAG 840
2Y 1423 TGGTATCGAAT 1433
Db 841 TGGTATCGAAT 851

RESULT 14
BM450315 1016 bp mRNA linear EST 05-FEB-2002
LOCUS AGNCOURT_6394296 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5494628
DEFINITION 5', mRNA sequence.
ACCESSION BM450315
VERSION BM450315.1 GI:18499355
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1016)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM12120 row: d column: 21
High quality sequence stop: 621.
Location/Qualifiers
1. 1016
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5494628"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SF0R6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

FEATURES
source

BASE COUNT 239 a 242 c 213 g 321 t 1 others
ORIGIN
Query Match 40.9%; Score 802.8; DB 12; Length 1016;
Best Local Similarity 98.9%; Pred No. 3.6e-183;
Matches 829; Conservative 0; Mismatches 7; Indels 2; Gaps 2;
2Y 704 CTCCTGGTTACCCATTCCTGTTGATTTGCTGTGTCATGAACACAGGTGAAGATTCT 763

11 CGCTGGTTTACCCATTGCTTTGATTTGCTGGTCATGAAAACACAGGTGAAGATTCT 70
764 TTCAATTGGTGGCTGATGGCCCATACATAATTTCTGGTGGCTGTGGTGGTCAATTGACAGCT 823
Db 71 TTCAATTGGTGGCTGATGGCCCATACATAATTTCTGGTGGCTGTGGTGGTCAATTGACAGCT 130
2Y 824 ACTATTATGGGAAGTTGGTGAATGACCACTCAACATTTGTTTGTATATATGTTCTTCTC 883
Db 131 ACTATTATGGGAAGTTGGTGAATGACCACTCAACATTTGTTTGTATATATATCTTCTC 190
2Y 884 CTCAATGACCTGATCTTTATGGTACAGAACCTGGTATTTCTTATTTATTTATTTATGATTTTC 943
Db 191 CTCAATGACCTGATCTTTATGGTACAGAACCTGGTATTTCTTATTTATTTATTTATGATTTTC 250
2Y 944 TGAATTTCAATGTAGCTTTGCTTTTGGCTCTCTAGTCTCTACCACTGACTTCTCTTATGG 1003
Db 251 TGAATTTCAATGTAGCTTTGCTTTTGGCTCTCTAGTCTCTACCACTGACTTCTCTTATGG 310
2Y 1004 AATACCTGCTGCAGAGATTTCAATGTTTCAAGATTTAGGCCACCCGTTATGGCTTACCTTGG 1063
Db 311 AATACCTGCTGCAGAGATTTCAATGTTTCAAGATTTAGGCCACCCGTTATGGCTTACCTTGG 370
2Y 1064 CTCCAATGTATATTGGTTTATATAATTTCTTATCCAGCTTCAAAAGAGGAGAGATTTC 1123
Db 371 CTCCAATGTATATTGGTTTATATAATTTCTTATCCAGCTTCAAAAGAGGAGAGATTTC 430
2Y 1124 TTTTCCCTGTGTATCCACTTATATGTTCTGTGGGCTGTGGCTCTCTCTGCACCTCAGA 1183
Db 431 TTTTCCCTGTGTATCCACTTATATGTTCTGTGGGCTGTGGCTCTCTCTGCACCTCAGA 490
2Y 1184 AATGTTACCACTTTGTTTTCACAGATATCGCTTGGAGCACTATCTGTGACATCGAATT 1243
Db 491 AATGTTACCACTTTGTTTTCACAGATATCGCTTGGAGCACTATCTGTGACATCGAATT 550
2Y 1244 GGCTGGCAATTAGGAACCTGCTTCTGTGTTGGGCTCTTGTGCAATTTCTCGCTGTGGCAC 1303
Db 551 GGCTGGCAATTAGGAACCTGCTTCTGTGTTGGGCTCTTGTGCAATTTCTCGCTGTGGCAC 610
2Y 1304 TGTTTCAGAGATATCACGGGCCCTTTGATTTGTATCCAGAAATTTACCGAATTTGCTACAG 1363
Db 611 TGTTTCAGAGATATCACGGGCCCTTTGATTTGTATCCAGAAATTTACCGAATTTGCTACAG 670
2Y 1364 ACCCAACCATCCACACTGTGCTCCAGAGGAGCACTGTGAAATGCTGTGTGGGAAAGAGT 1423
Db 671 ACCCAACCATCCACACTGTGCTCCAGAGGAGCACTGTGAAATGCTGTGTGGGAAAGAGT 730
2Y 1424 GGTTATGCAATTTCCAGCAGCTTCCCTTCTTCTCCAGAAATTTGAGGAGCTTCACTTCCAT 1483
Db 731 GGTTATGCAATTTCCAGCAGCTTCCCTTCTTCTCCAGAAATTTGAGGAGCTTCACTTCCAT 790
2Y 1484 -CAGAGTTCAGAGTCAAGTTACC-AAAACTTTTGCAGAGGAGCTTCTGGCCACCCGG 1539
Db 791 CCAGAGTTCAGAGTCAAGTTACC-AAAACTTTTGCAGAGGAGCTTCTGGCCACCCGG 848
RESULT 15
BUI92994
LOCUS BUI92994
DEFINITION AGNCOURT_7888267 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6141418
5', mRNA sequence.
ACCESSION BUI92994
VERSION BUI92994.1 GI:22706978
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 856)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM13461 row: j column: 11
High quality sequence stop: 725.
Location/Qualifiers
1..856
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6141418"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ASE COUNT 184 a 188 C 189 G 295 T
RIGIN

Query Match 40.4%; Score 791.6; DB 13; Length 856;
Best Local Similarity 98.1%; Pred. No. 1.7e-180;
Matches 822; Conservative 0; Mismatches 14; Indels 2; Gaps 2;
Y 587 GTAGCTTCGTATGTACACTACGTTGATAGCCATGACCTGGATGGTATATGACCAAGACTT 646
b 1 GTAGCTTCGTATGTACACTACGTTGATAGCCATGACCTGGATGGTATATGACCAAGACTT 60
Y 647 CCATTGCTGCTGGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTGAGTCGACCTC 706
b 61 CCATTGCTGCTGGAGTAGCAGCTGGGGCTATCTTAGGCTGGCCATTGAGTCGACCTC 120
Y 707 TTGGTTTACCATGCTTTGATTGCTGCTCATCAACACAGGTGGAGAGTTCTTTTC 766
b 121 TTGGTTTACCATGCTTTGATTGCTGCTCATCAACACAGGTGGAGAGTTCTTTTC 180
Y 767 ATTGGTTCGTGATGGCCCTCATATATTCTTGGTCCCTGCTGGTGTCTATTGACAGCTACT 826
b 181 ATTGGTTCGTGATGGCCCTCATATATTCTTGGTCCCTGCTGGTGTCTATTGACAGCTACT 240
Y 827 ATTATGGGAAGTTGGTATGTCACCACTCAACATTTGTTTATATGCTTTTACTCTC 886
b 241 ATTATGGGAAGTTGGTATGTCACCACTCAACATTTGTTTATATGCTTTTACTCTC 300
Y 887 ATGGACCTGATCTTTATGGTACAGAACCTGGTATTTCTATTATTAATGATTTCTGA 946
b 301 ATGGACCTGATCTTTATGGTACAGAACCTGGTATTTCTATTATTAATGATTTCTGA 360
Y 947 ATTTCATGTAGCTTTGCTTTGGCTCTCTAGTCTACCACTGACTTCTCTTATGGAT 1006
b 361 ATTTCATGTAGCTTTGCTTTGGCTCTCTAGTCTACCACTGACTTCTCTTATGGAT 420
Y 1007 ACCTGCTGCAGAGATTTGATGTTTCAAGATTTAGGCCACCCGGTATTGGCTTACCTTGGCTC 1066
b 421 ACCTGCTGCAGAGATTTGATGTTTCAAGATTTAGGCCACCCGGTATTGGCTTACCTTGGCTC 480
Y 1067 CAATGTATATTGGTTTATATTTCTTCTATCAGCCCTCACAAGAGGAGAGATTTCTTT 1126
b 481 CAATGTATATTGGTTTATATTTCTTCTATCAGCCCTCACAAGAGGAGAGATTTCTTT 540
Y 1127 TCCTGTGTATCCACTTATATGCTCTGTGGGCTGTGGCTCTCTCTGCACTTCAGAAAT 1186
b 541 TCCTGTGTATCCACTTATATGCTCTGTGGGCTGTGGCTCTCTCTGCACTTCAGAAAT 600
Y 1187 GTTACCACTTTGTTTCAACAGTATCGCTGGAGCACTATATCTGTGACATCGAATTGGC 1246
b 601 GTTACCACTTTGTTTCAACAGTATCGCTGGAGCACTATATCTGTGACATCGAATTGGC 660

QY 1247 TGCATTAGGAACGTCTTCTCTGTTGGGCTCTTGTCAATTTTCTGCTCTGTCGCACTGT 1306
Db 661 TGCATTAGGAACGTCTTCTCTGTTGGGCTCTTGTCAATTTTCTGCTCTGTCGCACTGT 720
QY 1307 TCAGAGGATATCAGGGG-CGCCTTGATTTGATATCCAGAA-TTTTACCGAATTGCTACAGA 1364
Db 721 TCAGAGGATATCAGGGGCCCCCTTGATTTGATATCCAGAAATTTTACCGAATTGCTACAGA 780
QY 1365 CCCAACCATCCACACTGTCCAGAGGCGACCTGTGAATGTCTGTGTGGGAAAGAG 1422
Db 781 ACCAACCATCCACACTGTCCAGAGGCGACCTGTGAATGTCTGTGTGGGAAAG 838

Search completed: January 13, 2004, 06:32:00
Job time : 4216 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	58.8	3.0	7218	1	US-08-232-463-14	Sequence 14, Appl
2	37.8	1.9	4403765	3	US-09-103-840A-2	Sequence 2, Appl
3	37.8	1.9	4411529	3	US-09-103-840A-1	Sequence 1, Appl
4	34.8	1.8	801	2	US-08-770-379-16	Sequence 16, Appl
5	34.8	1.8	801	3	US-08-757-669A-16	Sequence 16, Appl
6	34.8	1.8	801	4	US-09-298-568-3	Sequence 3, Appl
7	34.8	1.8	801	4	US-09-230-371A-16	Sequence 16, Appl
8	34.8	1.8	884	2	US-08-901-200A-11	Sequence 11, Appl
9	34.8	1.8	884	3	US-09-210-391-11	Sequence 11, Appl
10	34.8	1.8	6418	1	US-08-480-528A-11	Sequence 11, Appl
11	34.8	1.8	6418	1	US-08-473-666-11	Sequence 11, Appl
12	34.8	1.8	6418	5	PCF-US93-10520-11	Sequence 11, Appl
13	34.6	1.8	5117	3	US-08-854-585-1	Sequence 1, Appl
14	34.6	1.8	5117	4	US-09-441-533-1	Sequence 1, Appl
15	34.6	1.8	5117	5	PCF-US95-05512-1	Sequence 1, Appl
16	34.2	1.7	305	4	US-09-313-294A-7029	Sequence 120, App
17	34.2	1.7	1338	4	US-09-461-325-120	Sequence 11074, A
18	33.8	1.7	1377	4	US-09-253-991A-11074	Sequence 11004, A
19	33.8	1.7	1407	4	US-09-252-991A-11004	Sequence 11004, A
20	33.6	1.7	600	4	US-09-252-991A-5737	Sequence 5737, Ap
21	33.6	1.7	1053	4	US-09-253-991A-1697	Sequence 1697, Ap
22	33.6	1.7	1152	4	US-09-253-991A-1478	Sequence 1478, Ap
23	33.6	1.7	1827	4	US-09-253-991A-5756	Sequence 5756, Ap
24	33.6	1.7	2136	4	US-09-252-991A-5688	Sequence 5688, Ap
25	33.6	1.7	2907	2	US-09-018-628-17	Sequence 17, Appl
26	33.6	1.7	2907	3	US-09-273-378-17	Sequence 17, Appl
27	33.6	1.7	2907	3	US-09-018-635-26	Sequence 26, Appl

```

Best Local Similarity 3.8%; Pred. No. 1.5e-07;
Matches 15; Conservative 226; Mismatches 153; Indels 0; Gaps 0;

Y 821 GCTACTATTATGGGAAGTGTGGTATGACACATCAACATTTGTTTGTATAATGCTTTA 880
b 1042 GCTCAGGTCGAGGAGCTTGCATATATATATATATATATATATATATATATATAT 1101
Y 881 CTCCTCAGGACCTGATCTTATGATGACAGACCTGATTTCTATTAATTAATGAT 940
b 1102 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1161
Y 941 TTCTGAATTCATAGCTTTGCTTTGGCTCCTAGTCTACCACTGACTCTCTTA 1000
b 1162 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1221
Y 1001 TGAATACCTGCTGACAGATTCATGTCAGATTTAGGCACCGCTATGCTTACCT 1060
b 1222 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1281
Y 1061 TGCTCCATGATATTTGGTTTATAATTTCTTCATCCAGCTCACAAGAGAGAGAT 1120
b 1282 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1341
Y 1121 TTCTTTTCCCTGTATACCATATATATGCTCTGTGGCGCTGTGCTCTCTGCACTTC 1180
b 1342 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1401
Y 1181 AGAATGTTACCACTTTGTTTCAACATATCG 1214
b 1402 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYG 1435

RESULT 2
S-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
S-09-103-840A-2

Query Match 1.9%; Score 37.8; DB 3; Length 4403765;
Best Local Similarity 57.0%; Pred. No. 38;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Y 11 TGGCTAGTCAGGGGCTCGCGAGCGCTGGAAGGCGCAGCGGCGGCGGATACGG 70
b 2258699 TGACGGGTGGAGGGGCGAGCGCCAGCGGCGGCGGCGGCGGCGGCGGCGG 2258640
Y 71 CCCCGGTCGGGACAGCTCGGAGCTGCTGGGCGAGCCGAGAGCGGCGGCGGCGG 130
b 2258639 GCTCGGCGATCCCGGATCTCGGCGAGCAGCGGCGGCGGCGGCGGCGG 2258580
Y 131 A 131
b 2258579 A 2258579

```

```

RESULT 3
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 1.9%; Score 37.8; DB 3; Length 4411529;
Best Local Similarity 57.0%; Pred. No. 38;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 11 TGGCTAGTCAGGGGCTCGCGAGCGCTGGAAGGCGCAGCGGCGGCGGATACGG 70
Db 2261400 TGACGGGTGGAGGGGCGAGCGCCAGCGGCGGCGGCGGCGGCGGCGG 2261341
QY 71 CCCCGGTCGGGACAGCTCGGAGCTGCTGGGCGAGCCGAGAGCGGCGGCGGCGG 130
Db 2261340 GCTCGGCGATCCCGGATCTCGGCGAGCAGCGGCGGCGGCGGCGGCGG 2261281
QY 131 A 131
Db 2261280 A 2261280

RESULT 4
US-08-770-379-16/c
Sequence 16, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
OTHER INFORMATION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

```



```

; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
JS-08-770-379-16

Query Match 1.8%; Score 34.8; DB 2; Length 801;
Best Local Similarity 53.7%; Pred. No. 1.3;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

2Y 2 GGCTTGCCATGGCTAGTCGAGGGGCTCGGCAGCGCTGCTGAAGGCGCAGCGGGGCCAGCAGTG 61
Db 143 GCCCGCGAGGAGCCGCGCGCGCCCTGCCCCCGCGGGCGGCGGCTCCGCGCGGG 84

2Y 62 GGGATACGCGCCCGCGCTGCGGACAAAGCTGCGGAGCTGCTGGGCGAGCGGCGCGGGCG 121
Db 83 GCCTGGCGGCGAGCGCGCGCGCCCTGCCCCCGCGGGGAGCGGCGGCGCGGCGGTGTGGGGG 24

2Y 122 GCGCGGAGCACCGG 135
Db 23 GCGCGGGCGCGGG 10

```

```

RESULT 5
JS-08-757-669A-16/c
; Sequence 16, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-757-669A-16

Query Match 1.8%; Score 34.8; DB 3; Length 801;
Best Local Similarity 53.7%; Pred. No. 1.3;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 2 GGCTTGCCATGGCTAGTCGAGGGGCTCGGCAGCGCTGCTGAAGGCGCAGCGGGGCCAGCAGTG 61
Db 143 GCCCGCGAGGAGCCGCGCGCGCCCTGCCCCCGCGGGCGGCGGCTCCGCGCGGG 84

Qy 62 GGGATACGCGCCCGCGCTGCGGACAAAGCTGCGGAGCTGCTGGGCGAGCGGCGCGGGCG 121
Db 83 GCCTGGCGGCGAGCGCGCGCGCCCTGCCCCCGCGGGGAGCGGCGGCGCGGCGGTGTGGGGG 24

Qy 122 GCGCGGAGCACCGG 135
Db 23 GCGCGGGCGCGGG 10

RESULT 6
US-09-298-568-3/c
; Sequence 3, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; EARLIER FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 801
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-3

Query Match 1.8%; Score 34.8; DB 4; Length 801;
Best Local Similarity 53.7%; Pred. No. 1.3;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 2 GGCTTGCCATGGCTAGTCGAGGGGCTCGGCAGCGCTGGAAGGCGCAGCGGGGCCAGCAGTG 61
Db 143 GCCCGCGAGGAGCCGCGCGCGCCCTGCCCCCGCGGGCGGCGGCTCCGCGCGGG 84

Qy 62 GGGATACGCGCCCGCGCTGCGGACAAAGCTGCGGAGCTGCTGGGCGAGCGGCGCGGGCG 121
Db 83 GCCTGGCGGCGAGCGCGCGCGCCCTGCCCCCGCGGGGAGCGGCGGCGCGGCGGTGTGGGGG 24

Qy 122 GCGCGGAGCACCGG 135
Db 23 GCGCGGGCGCGGG 10

```

```

RESULT 7
US-09-230-371A-16/c
; Sequence 16, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S

```

APPLICANT: Moore, Patrick S
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
 TITLE OF INVENTION: USES THEREOF
 FILE REFERENCE: 45185-G-PCT-US
 CURRENT APPLICATION NUMBER: US/09/230,371A
 CURRENT FILING DATE: 1999-11-17
 PRIOR APPLICATION NUMBER: PCT/US97/13346
 PRIOR FILING DATE: 1997-07-22
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 16
 LENGTH: 801
 TYPE: DNA

ORGANISM: Kaposi's sarcoma-associated herpesvirus
 S-09-230-371A-16

Query Match 1.8%; Score 34.8; DB 4; Length 801;
 Best Local Similarity 53.7%; Pred. No. 1.3;
 Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
 Y 2 GGCTTCCCATGCTAGTCGAGGGGCTCGGAGCGGCTGAAGGGCAGCGGGGCCAGCAGTG 61
 b 143 GCCCGCGAGGAGCGCGGCGGCGGCTGCCCCCGGCGGCGGCTCCGGCCGG 84
 Y 62 GGATACGCCCGCGGCTGCGGACAAGCTGCGGAGCTCTCTGGGAGCCGAGAGCGGGCG 121
 b 83 GCCTGCGGCGAGCGCGGCGGAGCGGAGCGGAGGGGAGGGGCGGCGGCTGTGGGGG 24
 Y 122 GCGCGGAGCACCGG 135
 b 23 GCGCGGGCGCGG 10

RESULT 8

S-08-901-200A-11/c
 Sequence 11, Application US/08901200A
 Patent No. 5854071
 GENERAL INFORMATION:
 APPLICANT: OPPERMANN, HERMANN
 APPLICANT: OZKAYNAK, ENGIN
 APPLICANT: KUBERASAMPATH, THANGAVEL
 APPLICANT: RUEGER, DAVID C.
 APPLICANT: PANG, ROY H.L.
 APPLICANT: COHEN, CHARLES M.
 TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
 STREET: 45 SOUTH STREET
 CITY: HOPKINTON
 STATE: MA
 COUNTRY: USA
 ZIP: 01748

COMPUTER READABLE FORM: disk
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/901,200A
 FILING DATE: 28-JUL-1997
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: MEYERS, THOMAS C.
 REGISTRATION NUMBER: 36,989
 REFERENCE/DOCKET NUMBER: CRP-076DV2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 435-9001
 TELEFAX: (508) 435-6951
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 884 base pairs
 TYPE: nucleic acid

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1..884
 OTHER INFORMATION: /note= "hop-2 genomic sequence"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1..837
 OTHER INFORMATION: /note= "EXON ONE"
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 884
 OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
 OTHER INFORMATION: POSITIONS 884 IN THIS SEQUENCE AND POSITION 1 IN SEQ ID NO 14"
 US-08-901-200A-11

Query Match 1.8%; Score 34.8; DB 2; Length 884;
 Best Local Similarity 52.0%; Pred. No. 1.4;
 Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
 QY 1 AGGCTTGCATGGCTAGTCGAGGGGCTCGGACGCGCTGAAGGGCAGCGGGGCCAGCACT 60
 Db 500 AGCGGGGGGCGCTCAGCTGGGGCGCTCAGCGGGCGGCGATCGGCTCCGCGCGGCCA 441
 QY 61 GGGGATACGGCCCGGCTGCGGACAAGCTCGGGAGCTGCTGGGAGCCGAGAGCGGGCG 120
 Db 440 GGGCTGGGGAGCGCCCGGCGGCAAGAGGCTGGGCTCGCGGGGCGGCGGGCGGGG 381
 QY 121 GCGCGGAGCACCGGAGCGGAGTTATCTGGG 150
 Db 380 GGGCGGAGCGGGCGCGGCTCAGCG 351

RESULT 9

US-09-219-391-11/c
 Sequence 11, Application US/09219391
 Patent No. 6153583
 GENERAL INFORMATION:
 APPLICANT: OPPERMANN, HERMANN
 APPLICANT: OZKAYNAK, ENGIN
 APPLICANT: KUBERASAMPATH, THANGAVEL
 APPLICANT: RUEGER, DAVID C.
 APPLICANT: PANG, ROY H.L.
 APPLICANT: COHEN, CHARLES M.
 TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
 STREET: 45 SOUTH STREET
 CITY: HOPKINTON
 STATE: MA
 COUNTRY: USA
 ZIP: 01748
 COMPUTER READABLE FORM: disk
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/219,391
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/901,200
 FILING DATE: 28-JUL-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: MEYERS, THOMAS C.
 REGISTRATION NUMBER: 36,989
 REFERENCE/DOCKET NUMBER: CRP-076DV2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (508) 435-9001

TELEPHONE: (508) 435-6951
TELEFAX: (508) 435-6951
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..884
OTHER INFORMATION: /note= "hop-2 genomic sequence"
FEATURE:
NAME/KEY: exon
LOCATION: 1..837
OTHER INFORMATION: /note= "EXON ONE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 884..885
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"
FEATURE:
NAME/KEY: exon
LOCATION: 1088..1277
OTHER INFORMATION: /note= "EXON TWO"
FEATURE:
NAME/KEY: exon
LOCATION: 1350..1814
OTHER INFORMATION: /note= "EXON THREE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1834..1835
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"
FEATURE:
NAME/KEY: exon
LOCATION: 1883..2077
OTHER INFORMATION: /note= "EXON FOUR"
FEATURE:
NAME/KEY: exon
LOCATION: 2902..2981
OTHER INFORMATION: /note= "EXON FIVE"
FEATURE:
NAME/KEY: exon
LOCATION: 3507..3617
OTHER INFORMATION: /note= "EXON SIX"
FEATURE:
NAME/KEY: exon
LOCATION: 6116..6361
OTHER INFORMATION: /note= "EXON SEVEN"
US-08-480-528A-11

Query Match 1.8%; Score 34.8; DB 3; Length 884;
Best Local Similarity 52.0%; Pred. No. 1.4;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 AGGCTTGCGATGGCTAGTCGAGGGGCTCGGACGCGCTGAAGGCGACGGGGCCAGCAGT 60
DB 500 AGGCGGGGGGCGCTCAGCTGGGGCGCTCAGCGGGCGCATCGGCTCCGGCGGACCCA 441
QY 61 GGGGATACGGCCCCGGCTGGCGACACAGCTGCGGAGCTGTGGGCGACCCGAGAGGCGGGC 120
DB 440 GGGCTTGGGAGCGCCCCGACGCGCAAGGAGGCTGGGCTCGGCGGGCGCGGGCGGGC 381
QY 121 GCGGGGAGCACCACCGACCGAGTTATCTGGG 150
DB 380 GGGCGGGAGCGGGCGACCGCGGCTCAGCG 351

RESULT 10
US-08-480-528A-11/c
Sequence 11, Application US/08480528A
Patent No. 5652118
GENERAL INFORMATION:
APPLICANT: OPPERMAN, HERMANN
APPLICANT: OZKAYNAK, ENGIN
APPLICANT: KUBERASAMPATH, THANGAVEL
APPLICANT: RUEGER, DAVID C.
APPLICANT: PANG, ROY H. L.
APPLICANT: COHEN, CHARLES M.
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,528A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FENTON Esq., GILLIAN M.
REGISTRATION NUMBER: 36,508
REFERENCE/DOCKET NUMBER: CRP-076FW

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7560
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 6418 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..6361
OTHER INFORMATION: /note= "hop-2 genomic sequence"
FEATURE:
NAME/KEY: exon
LOCATION: 1..837
OTHER INFORMATION: /note= "EXON ONE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 884..885
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"
FEATURE:
NAME/KEY: exon
LOCATION: 1088..1277
OTHER INFORMATION: /note= "EXON TWO"
FEATURE:
NAME/KEY: exon
LOCATION: 1350..1814
OTHER INFORMATION: /note= "EXON THREE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1834..1835
OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"
FEATURE:
NAME/KEY: exon
LOCATION: 1883..2077
OTHER INFORMATION: /note= "EXON FOUR"
FEATURE:
NAME/KEY: exon
LOCATION: 2902..2981
OTHER INFORMATION: /note= "EXON FIVE"
FEATURE:
NAME/KEY: exon
LOCATION: 3507..3617
OTHER INFORMATION: /note= "EXON SIX"
FEATURE:
NAME/KEY: exon
LOCATION: 6116..6361
OTHER INFORMATION: /note= "EXON SEVEN"
US-08-480-528A-11

Query Match 1.8%; Score 34.8; DB 1; Length 6418;
Best Local Similarity 52.0%; Pred. No. 5.4;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 AGGCTTGCGATGGCTAGTCGAGGGGCTCGGACGCGCTGAAGGCGACGGGGCCAGCAGT 60
DB 500 AGGCGGGGGGCGCTCAGCTGGGGCGCTCAGCGGGCGCATCGGCTCCGGCGGACCCA 441
QY 61 GGGGATACGGCCCCGGCTGGCGACACAGCTGCGGAGCTGTGGGCGACCCGAGAGGCGGGC 120
DB 440 GGGCTTGGGAGCGCCCCGACGCGCAAGGAGGCTGGGCTCGGCGGGCGCGGGCGGGC 381
QY 121 GCGGGGAGCACCACCGACCGAGTTATCTGGG 150
DB 380 GGGCGGGAGCGGGCGACCGCGGCTCAGCG 351

RESULT 11
US-08-479-666-11/c

```

LOCATION: 2902..2981
OTHER INFORMATION: /note= "EXON FIVE"
FEATURE:
NAME/KEY: exon
LOCATION: 3507...3617
OTHER INFORMATION: /note= "EXON SIX"
FEATURE:
NAME/KEY: exon
LOCATION: 6116..6361
OTHER INFORMATION: /note= "EXON SEVEN"
US-08-479-666-11

Query Match 1.8%; Score 34.8; DB 1; Length 6418;
Best Local Similarity 52.0%; Pred. No. 5.4;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 AGGCTTGCCATGGCTACTGTCAGGGGCTCGCGACGGCTCAAGGGCAGCGGGGCCACGACGT 60
DB 500 AGGCGGGGGCGTCACTGGGGCGCTCAGCGGGCGCGCATCGGTCCTCGGGGCCGACCCA 441
QY 61 GGGGNATAGCGCCCGGTCGGTGGGACAGCTGCGGGAGCTGCTGGGACGCGAGAGCGGGC 120
DB 440 GGGCTGTGGGACGCCCGCGGCAAGAGAGCTGGGTCGGCGGGCGGGCGGGCGGGC 381
QY 121 GCGCGGGAGCACCGGACCGGAGTTCCTGGG 150
DB 380 GGGGCGGACGGGCGACCGCGGCTCAGCG 351

RESULT 12
PCT-US93-10520-11/c
Sequence 11, Application PC/TUS9310520
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: CREATIVE BIOMOLECULES, INC.
STREET: 45 SOUTH STREET
CITY: HOPKINTON
STATE: MA
COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10520
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/667,274
FILING DATE: 11-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,764
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/753,059
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,857
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/923,780
FILING DATE: 31-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,813
FILING DATE: 31-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829

```


RESULT 15
 RT-US95-05512-1
 Sequence 1, Application PC/TUS9505512
 GENERAL INFORMATION:
 APPLICANT: Tonks, Nicholas K. and stman, Arne
 TITLE OF INVENTION: Density Enhanced Protein Tyrosine
 TITLE OF INVENTION: Phosphatase
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSES: Marshall, O'Toole, Gerstein, Murray &
 ADDRESSES: Borun
 STREET: 233 South Wacker Drive, Suite 6300
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/05512

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM nucleic - nucleic search, using sw model

Run on: January 14, 2004, 13:55:05 ; Search time 658 Seconds
(without alignments)
10504.652 Million cell updates/sec

Title: US-09-922-225A-1
Perfect score: 1961
Sequence: 1 agccttgcatggtagtcg.....aggcttttgcatagaataa 1961

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1961	100.0	1961	11	US-09-922-225A-1
2	1017	51.9	1017	11	US-09-922-225A-7
3	935.4	47.7	1225	11	US-09-922-225A-5
4	897.6	45.8	953	15	US-10-050-704-35
5	580.4	29.6	880	11	US-09-922-225A-3
6	444.2	22.7	530	11	US-09-918-995-22803
7	371	18.9	372	9	US-09-777-564-39
8	371	18.9	372	15	US-10-015-219-39
9	291	14.8	445	15	US-10-102-524-980
10	237	12.1	500	13	US-10-029-386-10276
11	122.4	6.2	124	13	US-10-029-386-23985
12	90	4.6	3317	15	US-10-128-714-173
13	90	4.6	4016	15	US-10-128-714-5173
14	60	3.1	60	13	US-09-908-975-8397
15	57.4	2.9	427	11	US-09-918-995-37352

16	50.6	2.6	51	11	US-09-922-225A-101	Sequence 101, App
17	50.6	2.6	51	11	US-09-922-225A-102	Sequence 102, App
18	50.6	2.6	51	11	US-09-922-225A-103	Sequence 103, App
19	43.2	2.2	5856	13	US-10-311-455-789	Sequence 789, App
20	42.2	2.2	671	15	US-10-184-644-346	Sequence 346, App
21	42.2	2.2	671	15	US-10-184-644-346	Sequence 346, App
22	41.2	2.1	1089	15	US-10-184-644-266	Sequence 266, App
23	41.2	2.1	1089	15	US-10-184-644-266	Sequence 266, App
24	40.8	2.1	25554	12	US-10-017-161-695	Sequence 695, App
25	40.8	2.1	84510	12	US-10-292-798-607	Sequence 607, App
26	39	2.0	1041	14	US-10-027-632-121553	Sequence 121553,
27	39	2.0	1041	14	US-10-027-632-121553	Sequence 121553,
28	38.8	2.0	1254	13	US-10-027-632-204193	Sequence 204193,
29	38.8	2.0	1254	13	US-10-027-632-204193	Sequence 204193,
30	38	1.9	775	13	US-10-140-472-120	Sequence 120, App
31	38	1.9	775	13	US-10-141-761-120	Sequence 120, App
32	38	1.9	775	13	US-10-142-885-120	Sequence 120, App
33	38	1.9	775	13	US-10-158-790-120	Sequence 120, App
34	38	1.9	775	13	US-10-137-871-120	Sequence 120, App
35	38	1.9	775	13	US-10-140-805-120	Sequence 120, App
36	38	1.9	775	13	US-10-140-864-120	Sequence 120, App
37	38	1.9	775	13	US-10-140-923-120	Sequence 120, App
38	38	1.9	775	13	US-10-141-756-120	Sequence 120, App
39	38	1.9	775	13	US-10-141-759-120	Sequence 120, App
40	38	1.9	775	15	US-10-123-155-120	Sequence 120, App
41	38	1.9	775	16	US-10-146-731-120	Sequence 120, App
42	37.8	1.9	18301	12	US-10-242-355-708	Sequence 708, App
43	37.4	1.9	426	15	US-10-156-761-190	Sequence 190, App
44	37.4	1.9	7006	13	US-10-341-434-221	Sequence 221, App
45	37.4	1.9	8733	13	US-10-144-198-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-922-225A-1
; Sequence 1, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; TITLE OF INVENTION: Associated with Bipolar Disorder
; FILE REFERENCE: P-EA 4572
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1961
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)...(1842)
US-09-922-225A-1

Query Match		100.0%	Score 1961;	DB 11;	Length 1961;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 1961;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	1	AGGCTTGCCATGGCTAGTCGAGGGGCTCGGAGCGCCCTGAAGGGCAGCGGGGCCAGCAGT	60		
Db	1	AGGCTTGCCATGGCTAGTCGAGGGGCTCGGAGCGCCCTGAAGGGCAGCGGGGCCAGCAGT	60		
Qy	61	GGGATACGGCCCGGCTCGGACAGCTCGGGAGCTGCGGCGAGCGCGGAGCGGGC	120		
Db	61	GGGATACGGCCCGGCTCGGACAGCTGCGGGAGCTGCGGCGAGCGCGGAGCGGGC	120		
Qy	121	GGCGCGGAGCACCGGACCGAGTTATCTGGGAACAAAGCAGGCAAGTCTGGGCACCTGAA	180		
Db	121	GGCGCGGAGCACCGGACCGAGTTATCTGGGAACAAAGCAGGCAAGTCTGGGCACCTGAA	180		

181 GGATCTACTGCTTTCAAGTGTCTGCTTTTCAAGAGGTTATGTGCTGCTCTCTCTGAGCAAC 240
 181 GGATCTACTGCTTTTCAAGTGTCTGCTTTTCAAGAGGTTATGTGCTGCTCTCTCTGAGCAAC 240
 241 ATCTCTGATCTGTGATGAACATTCACATCTGAGGAGCCAAACACATCTCATCTATAGG 300
 241 ATCTCTGATCTGTGATGAACATTCACATCTGAGGAGCCAAACACATCTCATCTATAGG 300
 301 GAAGGGTTTCAAGTGTGGAATATTTCCACAGATATGCCATTCGCTCTCTATGCTTACCTG 360
 301 GAAGGGTTTCAAGTGTGGAATATTTCCACAGATATGCCATTCGCTCTCTATGCTTACCTG 360
 361 TTGCTTCAATGCTGAGGAGGAGTGTGATTCATGCAAGAAATTCACAACTAATAAGATCTT 420
 361 TTGCTTCAATGCTGAGGAGGAGTGTGATTCATGCAAGAAATTCACAACTAATAAGATCTT 420
 421 GTGTTTACTTTTTCGATGCTCTCTGCTGCTTTTCTGAGCTGTATTTCTGAACTTTACTTT 480
 421 GTGTTTACTTTTTCGATGCTCTCTGCTGCTTTTCTGAGCTGTATTTCTGAACTTTACTTT 480
 481 TACAAGGCTGTGCAAGAGTGTGGGTTGCACTGAGTGTGAATGAGTGTGCTGCTG 540
 481 TACAAGGCTGTGCAAGAGTGTGGGTTGCACTGAGTGTGAATGAGTGTGCTGCTG 540
 541 GTTCTCAGCACTGGCATGTTTCTCATCATCAGCATTCCTTCTGAGTGTGCTGCTG 600
 541 GTTCTCAGCACTGGCATGTTTCTCATCATCAGCATTCCTTCTGAGTGTGCTGCTG 600
 601 TACACTACGTTGATGACCATGCTGATGATGATGCAAGAGTTCATGCTGCTGCTG 660
 601 TACACTACGTTGATGACCATGCTGATGATGATGCAAGAGTTCATGCTGCTGCTG 660
 661 GGATGAGCACTGGGCTATCTTGGCTGGGCTTCTGAGTGTGCTGCTGCTGCTGCTG 720
 661 GGATGAGCACTGGGCTATCTTGGCTGGGCTTCTGAGTGTGCTGCTGCTGCTGCTG 720
 721 GCCTTTGATTTGCTGCTCATGAACACAGGTGGAGAGTTCCTTCTGATGCTGCTGCTG 780
 721 GCCTTTGATTTGCTGCTCATGAACACAGGTGGAGAGTTCCTTCTGATGCTGCTGCTG 780
 781 GCCTCATACTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 781 GCCTCATACTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 841 GTGATTTGCCACCACTCAACATTTGTTGATGAATGCTTTTCTGCTGCTGCTGCTGCTG 900
 841 GTGATTTGCCACCACTCAACATTTGTTGATGAATGCTTTTCTGCTGCTGCTGCTGCTG 900
 901 TATGTTACAGAACCTGTTATTTCTATTTAATGAATGCTTTTCTGCTGCTGCTGCTGCTG 960
 901 TATGTTACAGAACCTGTTATTTCTATTTAATGAATGCTTTTCTGCTGCTGCTGCTGCTG 960
 961 TTTGCTTTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 961 TTTGCTTTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 1021 TTTTATGTTTCAAGTGTGGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 1021 TTTTATGTTTCAAGTGTGGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 1081 TTTTATGTTTCTTCTATCCAGCTCACAAGAGGAGAGATTTCTTCTGCTGCTGCTGCTGCT 1140
 1081 TTTTATGTTTCTTCTATCCAGCTCACAAGAGGAGAGATTTCTTCTGCTGCTGCTGCTGCT 1140
 1141 CTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 1141 CTATATGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 1201 TTTTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 1201 TTTTCAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 1261 GTCTTCTGTTGGGCTCTTGTGCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320

1261 GTCTTCTGTTGGGCTCTTGTGCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 1321 GGCCCTCTGATTTGATCCAGAAATTTTACCAATTTTACCAATTTTACCAATTTTACCAATTT 1380
 1321 GGCCCTCTGATTTGATCCAGAAATTTTACCAATTTTACCAATTTTACCAATTTTACCAATTT 1380
 1381 GTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 1381 GTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
 1441 AGCTTCTCTTCTTCTGACAAATTCGAGCTTTCAGTTTTCATTCATTCATTCATTCATTCATTC 1500
 1441 AGCTTCTCTTCTTCTGACAAATTCGAGCTTTCAGTTTTCATTCATTCATTCATTCATTCATTC 1500
 1501 TTACCAAAACCTTTTTCAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 1501 TTACCAAAACCTTTTTCAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
 1561 GACCAAGATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 1561 GACCAAGATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
 1621 GATTTGAGACCATTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 1621 GATTTGAGACCATTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
 1681 TGGATCAGCTTGGCTTATAGACCATTTCTTGTGCTTCTGATGCTTCTGATGCTTCTGATGCTTCT 1740
 1681 TGGATCAGCTTGGCTTATAGACCATTTCTTGTGCTTCTGATGCTTCTGATGCTTCTGATGCTTCT 1740
 1741 GCATTTCTATGTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
 1741 GCATTTCTATGTCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
 1801 AAACCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 1801 AAACCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
 1861 CCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
 1861 CCCAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
 1921 GGCCTGTAACTTTGTAATAAGGCTTCTGACATGAAAGAA 1961
 1921 GGCCTGTAACTTTGTAATAAGGCTTCTGACATGAAAGAA 1961

RESULT 2

US-09-922-225a-7
 ; Sequence 7, Application US/09922225A
 ; Publication NO. US20030104385A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Evans, Glen A.
 ; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
 ; TITLE OF INVENTION: Associated with Bipolar Disorder
 ; FILE REFERENCE: P-EA 4672
 ; CURRENT APPLICATION NUMBER: US/09/922,225A
 ; NUMBER OF SEQ ID NOS: 117
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 1017
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1017)
 US-09-922-225a-7

Query Match 51.9%; Score 1017; DB 11; Length 1017;
 Best Local Similarity 100.0%; Pred. No. 4,9e-303;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

10 ATGGCTAGTCGAGGGCTCGGACGCTGAGGGGAGCGGGGCGGAGCGGATACG 69
Db |
1 ATGGCTAGTCGAGGGCTCGGACGCTGAGGGGAGCGGGGCGGAGCGGATACG 60
Qy |
70 GCCCGGCTGCGGACAAAGCTCGGAGAGCTGCTGGGACGCGGAGCGGCGGAG 129
Db |
61 GCCCGGCTGCGGACAAAGCTCGGAGAGCTGCTGGGACGCGGAGCGGCGGAG 120
Qy |
130 CACGGACCGAGTTATCTGGGACAAAGCAGGACAGCTGCGGACCTGAGAGTCTACT 189
Db |
131 CACGGACCGAGTTATCTGGGACAAAGCAGGACAGCTGCGGACCTGAGAGTCTACT 180
Qy |
190 GCTTTCAAGTCTGCTCTTTACGACAGGTTATGCTGCTCTCTCTGAGCAACATCTCTGAC 249
Db |
181 GCTTTCAAGTCTGCTCTTTACGACAGGTTATGCTGCTCTCTCTGAGCAACATCTCTGAC 240
Qy |
250 TGTGATGAAACATTTCACTACTGCGGACGACACACACTACCTCATCTATGCGGAGGTTT 309
Db |
241 TGTGATGAAACATTTCACTACTGCGGACGACACACACTACCTCATCTATGCGGAGGTTT 300
Qy |
310 CAGACTTGGGAAATTTCCCGACGATATGCCATTCGCTCTCTATGCTTACCTGTTGCTTCAT 369
Db |
301 CAGACTTGGGAAATTTCCCGACGATATGCCATTCGCTCTCTATGCTTACCTGTTGCTTCAT 360
Qy |
370 GCCTGGCCAGCTGCAATTCATGCAAGAAATTTCTACAACTAATAGATTTCTGTGTTTTAC 429
Db |
361 GCCTGGCCAGCTGCAATTCATGCAAGAAATTTCTACAACTAATAGATTTCTGTGTTTTAC 420
Qy |
430 TTTTGTGCAATCTCTCTGCTTTTGTGAGCTGTTATTTGTGAACTTTACTTTTCAAGGCT 489
Db |
421 TTTTGTGCAATCTCTCTGCTTTTGTGAGCTGTTATTTGTGAACTTTACTTTTCAAGGCT 480
Qy |
490 GTGTGCAAGAAATTTGGGTTGCGAGTGGGAAATGATGCTAGCCTTTGTTGTTCTCAGC 549
Db |
481 GTGTGCAAGAAATTTGGGTTGCGAGTGGGAAATGATGCTAGCCTTTGTTGTTCTCAGC 540
Qy |
550 ACTGGCAGTTTGTCTCATCATCAGCAATCTCTCTAGTAGTCTCTGATGATACACTAG 609
Db |
541 ACTGGCAGTTTGTCTCATCATCAGCAATCTCTCTAGTAGTCTCTGATGATACACTAG 600
Qy |
610 TTGATAGCCATGACTGGATGTTATGCAAGAAATTTCCATTCGCTGCTGGGAGTAGCA 669
Db |
601 TTGATAGCCATGACTGGATGTTATGCAAGAAATTTCCATTCGCTGCTGGGAGTAGCA 660
Qy |
670 GCTGGGCTATCTTAGCTGGCATTGAGTGGCATTGAGTGGCATTGAGTGGCATTGAGT 729
Db |
661 GCTGGGCTATCTTAGCTGGCATTGAGTGGCATTGAGTGGCATTGAGTGGCATTGAGT 720
Qy |
730 TTGCTGGTCAATGAAACACAGGTGGAGAGTTTCTTTCATTCGCTGATGGCCCTCAT 789
Db |
721 TTGCTGGTCAATGAAACACAGGTGGAGAGTTTCTTTCATTCGCTGATGGCCCTCAT 780
Qy |
790 CTATTTCTGGTGGCTGGTGGTCAATGACAGCTACTATTTATGGAGTTGGTGGATGCA 849
Db |
781 CTATTTCTGGTGGCTGGTGGTCAATGACAGCTACTATTTATGGAGTTGGTGGATGCA 840
Qy |
850 CCACCTCAACATTTGTTGTAATATGCTTTTACTCCTCATGGAACCTGATCTTTATGTACA 909
Db |
841 CCACCTCAACATTTGTTGTAATATGCTTTTACTCCTCATGGAACCTGATCTTTATGTACA 900
Qy |
910 GAAACCTGGTATTTCTATTTAATGGAATTTCTGAAATTTCAATGAGCTTTGCTTTG 969
Db |
901 GAAACCTGGTATTTCTATTTAATGGAATTTCTGAAATTTCAATGAGCTTTGCTTTG 960
Qy |
970 GCTCTCTAGTCCCTACCACTGCTCTCTTATGGAATACCTGCTGAGAGATTTTCAT 1026
Db |
961 GCTCTCTAGTCCCTACCACTGCTCTCTTATGGAATACCTGCTGAGAGATTTTCAT 1017

RESULT 3
US-09-922-225a-5
; Sequence 5, Application US/09922225A

Publication No. US20030104385A1
GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; FILE OF INVENTION: Associated with Bipolar Disorder
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1225
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (114)...(1106)
US-09-922-225a-5

Query Match 47.7%; Score 935.4; DB 11; Length 1225;
Best Local Similarity 99.4%; Pred. No. 9.4e-278;
Matches 939; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1017 GAGATTTTCATGTTTCAGAAATTTAGGCCACCCGTTATGGCTTACCTTGGCTTCCCAATCTATAT 1076
Db 281 GTGTGTCAGTTTCAGAAATTTAGGCCACCCGTTATGGCTTACCTTGGCTTCCCAATCTATAT 340
Qy 1077 TTGTTTATATAATTTTCTTCATCCAGCTTCACAAAGAGGAGAGATTTCTTTTCCCTGTGTA 1136
Db 341 TTGTTTATATAATTTTCTTCATCCAGCTTCACAAAGAGGAGAGATTTCTTTTCCCTGTGTA 400
Qy 1137 TCACATTTATATGCTCTGTGGGCTCTGTGCTCTCTGCACTTCAGAAATGTTACACTT 1196
Db 401 TCACATTTATATGCTCTGTGGGCTCTGTGCTCTCTGCACTTCAGAAATGTTACACTT 460
Qy 1197 TGTGTTTCAACGATATCGCTTGAGCACTATATCTGTGACATCGAATTTGCTGGCAATGAG 1256
Db 461 TGTGTTTCAACGATATCGCTTGAGCACTATATCTGTGACATCGAATTTGCTGGCAATGAG 520
Qy 1257 AACTGTCTTCTCTTTGGGCTCTGTGCTCTCTGCTGTGGCACTTTCAGAGGATA 1316
Db 521 AACTGTCTTCTCTTTGGGCTCTGTGCTCTCTGCTGTGGCACTTTCAGAGGATA 580
Qy 1317 TCACGGGCCCCCTTGATTTGTATCCAGAAATTTTACCGAAATTTGCTACAGACCCCAACCATCA 1376
Db 581 TCACGGGCCCCCTTGATTTGTATCCAGAAATTTTACCGAAATTTGCTACAGACCCCAACCATCA 640
Qy 1377 CACTGTCCCAAGAGGAGAGCACTGTGTAATGCTGTGTGGGAAAAGAGTGTATCGATTTC 1436
Db 641 CACTGTCCCAAGAGGAGAGCACTGTGTAATGCTGTGTGGGAAAAGAGTGTATCGATTTC 700
Qy 1437 CAGCAGCTTCTCTTCTTCCCTGACAAATTTGGCAGCTTCAGTTTCATTCATCCATCAGAGTT 1496
Db 701 CAGCAGCTTCTCTTCTTCCCTGACAAATTTGGCAGCTTCAGTTTCATTCATCCATCAGAGTT 760
Qy 1497 TCAGTTACAAAACCTTTTTCAGAGAGCACTGTGGCACCAGGATTTCTTCTTCTTCTTCT 1556
Db 761 TCAGTTACAAAACCTTTTTCAGAGAGCACTGTGGCACCAGGATTTCTTCTTCTTCTTCT 820
Qy 1557 GAATGACCAAGTCTAGAGAGCCATCCAGATATATTTGATATTCAGTAAATGCCATTATTT 1616
Db 821 GAATGACCAAGTCTAGAGAGCCATCCAGATATATTTGATATTCAGTAAATGCCATTATTT 880
Qy 1617 AGTGGATTTGCAACCATGAGAGAAACACCCCGGGAGCCAAAATATTCATCCAAATAAGA 1676
Db 881 AGTGGATTTGCAACCATGAGAGAAACACCCCGGGAGCCAAAATATTCATCCAAATAAGA 940
Qy 1677 AGATGATCAGCTTGGCTTATAGACCATTTCTTGTGATGCTTCTAGATCTTCAAGCTGCT 1736
Db 941 AGATGATCAGCTTGGCTTATAGACCATTTCTTGTGATGCTTCTAGATCTTCAAGCTGCT 1000
Qy 1737 GCGGCGATTTCTATGTCCTTCTTCTGTCAGATCAGTATACAGTGTACGTAACCTACACCAT 1796


```

460 TGTGTTCAACGATATCGCTGGAGCACTATATCTGTGACATCCAAATGGCTGGCATTAGG 519
1257 AACTGCTTCTCTGTTGGGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTCAGAGGATA 1316
520 AACTGCTTCTCTGTTGGGCTCTTGTCAATTTCTGCTCTGTGGCACTGTTTCAGAGGATA 579
1317 TCAGGGGCCCTCTGATTTGTATCCAGAAATTTACCGAATTTGTACAGACCCCAACATCCA 1376
580 TCAGGGGCCCTCTGATTTGTATCCAGAAATTTACCGAATTTGTACAGACCCCAACATCCA 639
1377 CACTGTCCAGAGGAGCACTCTGATGTCGTGTGGGAAGAGTGTATCGATTTCC 1436
640 CACTGTCCAGAGGAGCACTCTGATGTCGTGTGGGAAGAGTGTATCGATTTCC 699
1437 CAGCAGCTTCTCTTCTCTGACAAATTTGGCAGCTTTCAGTTCATTTCCATCAGAGTTCCAGAGG 1496
700 CAGCAGCTTCTCTTCTCTGACAAATTTGGCAGCTTTCAGTTCATTTCCATCAGAGTTCCAGAGG 759
1497 TCAGTTACCAACACCTTTTCAGAGGAGCTCTGGCCAGCCCGATTTCTTCTACTGACAT 1556
760 TCAGTTACCAACACCTTTTCAGAGGAGCTCTGGCCAGCCCGATTTCTTCTACTGACAT 819
1557 GAATGACCAAGATCTAGAGAGCCATCCAGATATATTGATATCAGTAAATGCCATTATT 1616
820 GAATGACCAAGATC-AGAAGAGCCATCCAGATATATTGATATCAGTAAATGCCATTATT 878
1617 AG 1618
879 AG 880

RESULT 6
US-09-918-995-22803
; Sequence 22803, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-755
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22803
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(530)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22803

Query Match 22.7%; Score 444.2; DB 11; Length 530;
Best Local Similarity 95.0%; Pred. No. 3.8e-126;
Matches 477; Conservative 0; Mismatches 4; Indels 21; Gaps 1;

2Y 1083 TATATTTTCTTCATCCAGCTCTCAAGAGGAGAGATTTCTTCTGCTGTATCCACT 1142
DB 28 TATATTTTCTTCATCCAGCTCTCAAGAGGAGAGATTTCTTCTGCTGTATCCACT 87
2Y 1143 TATATGTCCTGTGGCGCTGTGGCTCTCTCTGCACTT-----CA 1181
DB 88 TATATGTCCTGTGGCGCTGTGGCTCTCTCTGCACTT-----CA 1181
2Y 1182 GAATGTTTACCACTTTGTGTTTCAACGATATCGCTGGAGCACTTATACGTGACATCGAA 1241
DB 148 GAATGTTTACCACTTTGTGTTTCAACGATATCGCTGGAGCACTTATACGTGACATCGAA 207
2Y 1242 TTGGCTGGCATTAGGAAGTGTCTTCTGTTTGGGCTCTTGTGCACTTTTCTCGCTCTGTGGC 1301

```

```

DB 208 TTGGCTGGCATTAGGAAGTGTCTTCTGTTTGGGCTCTTGTCAATTTCTCGCTCTGTGGC 267
QY 1302 ACTGTTTCAGAGGATATCAGGGCCCTTGTGATTTGTATCCAGAAATTTTACCGAATTTGCTAC 1361
DB 268 ACTGTTTCAGAGGATATCAGGGCCCTTGTGATTTGTATCCAGAAATTTTACCGAATTTGCTAC 327
QY 1362 AGACCCAAACATCCACACTGTCCAGAAAGGAGAGCTGTGAATGTCTGTGTGGGAAAGA 1421
DB 328 AGACCCAAACATCCACACTGTCCAGAAAGGAGAGCTGTGAATGTCTGTGTGGGAAAGA 387
QY 1422 GTGGTATCCATTTCCAGCAGCTTCTTCTTCTGACAAATTTGGCAGCTTTCAGTTCATTTCC 1481
DB 388 GTGGTATCCATTTCCAGCAGCTTCTTCTTCTGACAAATTTGGCAGCTTTCAGTTCATTTCC 447
QY 1482 ATCAGAGTTTCAGAGGTCAGTTACCAAAACCTTTTTCAGAAAGAGCTCTGGCCACCCGGAT 1541
DB 448 ATCAGAGTTTCAGAGGTCAGTTACCAAAACCTTTTTCAGAAAGAGCTCTGGCCACCCGGGA 507
QY 1542 TGTTCCTACTGACATGAATGAC 1563
DB 508 GGTTCCTACTGACATGAATGAC 529

RESULT 7
US-09-777-564-39/c
; Sequence 39, Application US/09777564
; Patent No. US2002022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INFECTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 39
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(372)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-39

Query Match 18.9%; Score 371; DB 9; Length 372;
Best Local Similarity 99.7%; Pred. No. 1.3e-103;
Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1532 CCACCCGGATTGTTCCTACTGACATGAATGACCAATCTAGAGAGCCATCCAGATATA 1591
DB 372 CCACCCGGATTGTTCCTACTGACATGAATGACCAATCTAGAGAGCCATCCAGATATA 313
QY 1592 TTGATATCAGTAATGCCATTTATGTCGATTTTGACACCATGAGAGAAACACCCCGG 1651
DB 312 TTGATATCAGTAATGCCATTTATGTCGATTTTGACACCATGAGAGAAACACCCCGG 253
QY 1652 AGCCAAATATTCATCCAATAAAGAGAAATGGATCAGCTTGGCCTATAGACCATTCCTTG 1711
DB 252 AGCCAAATATTCATCCAATAAAGAGAAATGGATCAGCTTGGCCTATAGACCATTCCTTG 193
QY 1712 ATGCTTCTAGATCTTCAAGCTGCTCGGGCATTTCTATGTCCTCCCTTCTGTCAGATCAGT 1771
DB 192 ATGCTTCTAGATCTTCAAGCTGCTCGGGCATTTCTATGTCCTCCCTTCTGTCAGATCAGT 133
QY 1772 ATACAGTGTACGTAATACTACCATCTCTCAAAACCCCGGAAAGCAAAATCAGGAAGA 1831
DB 132 ATACAGTGTACGTAATACTACCATCTCTCAAAACCCCGGAAAGCAAAATCAGGAAGA 73
QY 1832 AAGTGGAGGTTAGCAACACACCTGTGGCCCAAGGACAAACCACTTTGTTTAACTATTGA 1891

```

b 72 AAAGTGGAGTTAGACACACACCTGTGGCCCAAGGACCAACCACTCTGTACTATTGA 13
y 1892 TTCCAGTGACCT 1903
b 12 TTCCAGTGACCT 1

RESULT 8

S-10-015-219-39/c
Sequence 39, Application US/10015219
Publication No. US2003008299A1
GENERAL INFORMATION:
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.493C1
CURRENT APPLICATION NUMBER: US/10/015,219
CURRENT FILING DATE: 2002-03-02
NUMBER OF SEQ ID NOS: 1739
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 354
OTHER INFORMATION: n = A,T,C or G
S-10-015-219-39

Query Match 18.9%; Score 371; DB 15; Length 372;
Best Local Similarity 99.7%; Pred. No. 1.3e-103;
Matches 371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

y 1532 CCACCCGGATTGTCCTACTGACATGAATGACCAATCTAGAAGAGCCATCCAGATATA 1591
b 372 CCACCCGGATTGTCCTANTGACATGAATGACCAATCTAGAAGAGCCATCCAGATATA 313
y 1592 TTGATATAGTAAGTAAAGCCATTATTAGTGGATTGGACCATGAGAAACACCCCGGG 1651
b 312 TTGATATAGTAAGTAAAGCCATTATTAGTGGATTGGACCATGAGAAACACCCCGGG 253
y 1652 AGCAAAATATTATCATCAATAAAGAGATGGATCAGCTTGCCCTATAGACCATTCCTTG 1711
b 252 AGCAAAATATTATCATCAATAAAGAGATGGATCAGCTTGCCCTATAGACCATTCCTTG 193
y 1712 ATGCTTCTAGATCTTCAAGCTGCTGGGGCAATCTATGTCCCTTCTGTGCATCAGT 1771
b 192 ATGCTTCTAGATCTTCAAGCTGCTGGGGCAATCTATGTCCCTTCTGTGCATCAGT 133
y 1772 ATACAGTGTAGCTAAACTACACCATCTCAAAACCCCGGAAAGCAAGCAATCAGGAAGA 1831
b 132 ATACAGTGTAGCTAAACTACACCATCTCAAAACCCCGGAAAGCAAGCAATCAGGAAGA 73
y 1832 AAAGTGGAGTTAGCAACACACCTGTGGCCCCCAAGCAACCATCTTTGTTAACTATTGA 1891
b 72 AAAGTGGAGTTAGCAACACACCTGTGGCCCCCAAGCAACCATCTTTGTTAACTATTGA 13

y 1892 TTCCAGTGACCT 1903

b 12 TTCCAGTGACCT 1

RESULT 9

S-10-102-524-980
Sequence 980, Application US/10102524
Publication No. US20030109434A1
GENERAL INFORMATION:
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.493C1
CURRENT APPLICATION NUMBER: US/10/015,219
CURRENT FILING DATE: 2002-03-02
NUMBER OF SEQ ID NOS: 1739
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 372
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 354
OTHER INFORMATION: n = A,T,C or G
S-10-015-219-39

APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
FILE REFERENCE: 210121.572
CURRENT APPLICATION NUMBER: US/10/102,524
CURRENT FILING DATE: 2002-03-19
NUMBER OF SEQ ID NOS: 1863
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 980
LENGTH: 445
TYPE: DNA
ORGANISM: Homo sapiens
US-10-102-524-980

Query Match 14.8%; Score 291; DB 15; Length 445;
Best Local Similarity 100.0%; Pred. No. 7.9e-79;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 380 CTGCATTTTCATGCAAGAAATCTACAAACTAATAGATTCTTGTTTACTTTTCGGAT 439
DB 155 CTGCATTTTCATGCAAGAAATCTACAAACTAATAGATTCTTGTTTACTTTTCGGAT 214
QY 440 GTCTTCTGGCTTTTGTGAGCTGTATTGTGAATCTTTTACATTTTACAAGGCTGTGCAAGA 499
DB 215 GTCCTCTGGCTTTTGTGAGCTGTATTGTGAATCTTTTACATTTTACAAGGCTGTGCAAGA 274
QY 500 AGTTTGGTTCACGTGAGTGAATGATGCTAGCTTCTTGTTCTCAGCACTGGCATGT 559
DB 275 AGTTTGGTTCACGTGAGTGAATGATGCTAGCTTCTTGTTCTCAGCACTGGCATGT 334
QY 560 TTGTCTCATCATCAGCATTCCTCTAGTAGCTTCTGTATGTACACTAGCTTGTATAGCCA 619
DB 335 TTGTCTCATCATCAGCATTCCTCTAGTAGCTTCTGTATGTACACTAGCTTGTATAGCCA 394
QY 620 TGACTGGATGTATATGACCAAGACTTCCATTTGCTGTGCTGGGAGTAGCAG 670
DB 395 TGACTGGATGTATATGACCAAGACTTCCATTTGCTGTGCTGGGAGTAGCAG 445

RESULT 10

US-10-029-386-10276/c
Sequence 10276, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David K.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
FILE REFERENCE: AECOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 10276
LENGTH: 500
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC000381.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.48
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
OTHER INFORMATION: NT HIT: g13376061, EVALUATE 1.00e-127
OTHER INFORMATION: SWISSPROT HIT: Q86677, EVALUATE 2.10e+00
OTHER INFORMATION: EST_HUMAN HIT: AA886018.1, EVALUATE 0.00e+00
US-10-029-386-10276

Query Match 12.1%; Score 237; DB 13; Length 500;
Best Local Similarity 98.0%; Pred. No. 4.2e-62;
Matches 240; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1716 TTCTAGATCTTCAAGCTGTGCGGGAATCTTATGTCCTTCTGTGATCATCATATAC 1775

Db 375 TTTGAGATCTTCAAGCTGTGGGGCATCTATGTCCCTCTGTGAGATGATAC 316
Qy 1776 AGTGTAGTAACTACACCATCTCTCAAAACCCCGAAGCAAGCAATCAGGAAGAAAG 1835
Db 315 AGTGTAGTAACTACACCATCTCTCAAAACCCCGAAGCAAGCAATCAGGAAGAAAG 256
Qy 1836 TGGAGGTTAGCAACACACACCTGTGGGCCCCCAAGGACCAACCATCTGTGTTAACTATTGATTCC 1895
Db 255 TGGAGTTAGCAACACACACCTGTGGGCCCCCAAGGACCAACCATCTGTGTTAACTATTGATTCC 196
Qy 1896 AGTGACCTGATCCCTGGAAGTATCCCTCTGTAACTTTGTAATAAGTCTTTGACAT 1955
Db 195 AGTGACCTGATCCCTGGAAGTATCCCTCTGTAACTTTGTAATAAGTCTTTGACAT 136
Qy 1956 GAAAA 1960
Db 135 GAATA 131

RESULT 11
US-10-029-386-23985/c
; Sequence 23985, Application US/10029386
; Publication No. US2003019404A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23985
; LENGTH: 124
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC000381.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.48
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
; OTHER INFORMATION: NT HIT: g114770432, EVALUATE 2.00e-61
; OTHER INFORMATION: EST_HUMAN HIT: BG191573.1, EVALUATE 3.00e-61
US-10-029-386-23985

Query Match 6.2%; Score 122.4; DB 13; Length 124;
Best Local Similarity 99.2%; Pred. No. 4.3e-27;
Matches 123; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1722 ATCTTCAAAGCTGTGGGGCATCTATGTCCCTCTGTGAGATCAGTATACAGTGA 1781
Db 124 ATCTTCAAAGCTGTGGGGCATCTATGTCCCTCTGTGAGATCAGTATACAGTGA 65
Qy 1782 CGTAACTACACCATCTCTCAAAACCCCGAAGCAAGCAATCAGGAAGAAAGTGGAGG 1841
Db 64 CGTAACTACACCATCTCTCAAAACCCCGAAGCAAGCAATCAGGAAGAAAGTGGAGG 5
Qy 1842 TTAG 1845
Db 4 TTAG 1

RESULT 12
US-10-128-714-173/c
; Sequence 173, Application US/10128714
; Publication No. US2003019013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M

; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 173
; LENGTH: 3317
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-173

Query Match 4.6%; Score 90; DB 15; Length 3317;
Best Local Similarity 51.1%; Pred. No. 4.5e-16;
Matches 272; Conservative 0; Mismatches 245; Indels 15; Gaps 2;

Qy 223 GCTGCTCTCTGAGCAACATCTCTGACTGTGATGAAACAATCAACTACTGGGAGCAACA 282
Db 534 GCTGGGCACTCGCTCCGATTCAAGACTCGGATGAAGTCTTCAATTTTGGGAGCCGCG 475
Qy 283 CACTACCTCATCTATGGGAAGGGTTTCAGACTTGGGAATATTCCTCCAGCATATGCCATT 342
Db 474 CACTACCTAGACCAAGGTTACGAGGTTACAAAGCTGGGAGTATCTCCGCGTATTCAATT 415
Qy 343 CGCTCTCTATGCTTACCTGTGCTTCAATGCTGCGCAGCTGCAATTTTCATGCAAGAATTCTTA 402
Db 414 CGAAGCTGGCTCTACGTTCTGCCCATGCTGGG---GTTGGCAAAATAGTGTGATTTTC 358
Qy 403 CAAACTAATPAAGATTTCTTGTTTACTTTTTCGATGTCCTTCTGCTTTTGTGAGCTGT 462
Db 357 TCTAGTGATAAGACATCCGGAATTTTATCTACTACGTTTCGTTCTGGCAGCAGCTCTGCCGCC 298
Qy 463 ATTTGTGAACCTTTACTTTTACAAGGCTGTGTGCAAGAGTTTGGGTTGCACTGAGTGA 522
Db 297 GCCTGGCAGAGAGACTATATCTCAGCAATCTGTGCGACATTTGAAACCCAGAGTCCGCCCTC 238
Qy 523 ATGATGCTAGCCTTCTGTTCTCAGCACTGGGATGTTTGTCTCATCATCAGCATTCCTT 582
Db 237 CTCTTCTGATCATTTGCTGCTTTTCACTCCGGGCAATGTTCCATGCGTCTGCGACATTTCTTA 178
Qy 583 CCTAGTAGCTTCTGTATGTACACTAGCTTGTAGCCATGAC-----TGATGG 630
Db 177 CCGTCTAGTTTACCATGTATATCTTCATGCTAGGCTTTGCTATCTTTTCTAGATTGGAGG 118
Qy 631 TATATGACAAAGATTTCCATGCTGTGCTGGGAGTAGCAGCTGGGGCTATCTTAGCTGG 690
Db 117 GGAGGCCAGAAAGACGGCGCAGGGTATTATGTGGTTCCGTTCTCGGAGCAATCATGGGCTGG 58
Qy 691 CCATTTCAGTGCAGCTTGGTTTACCCATTTGCTTTGATTTTGTCTGCTCATGA 742
Db 57 CCATTTCGGAGCTCTTATCATTTCTTCTTCTTGGAGAGTCCGATCA 6

RESULT 13
US-10-128-714-5173/c
; Sequence 5173, Application US/10128714
; Publication No. US2003019013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 05:21:49 ; Search time 18 seconds

(without alignments)
1596.295 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKGSASSGDT.....VNYTLKPRKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1064	32.5	603	1 YTH3 CAEEL	P54002 Caenorhabdi
2	973	23.7	577	1 YFV5 SCHPO	Q9P7Q9 schizosacch
3	736	22.5	555	1 ALG9 YEAST	P53868 saccharomyc
4	214.5	6.6	616	1 YG02 YEAST	P30777 saccharomyc
5	144.5	4.4	516	1 SMP3 YEAST	Q04174 saccharomyc
6	128.5	3.9	551	1 YN81 YEAST	P53730 saccharomyc
7	123.5	3.8	533	1 YADC SCHPO	Q09837 schizosacch
8	116	3.5	662	1 CYOB BUCAL	P57543 buchnera ap
9	115	3.5	386	1 BSGY ACEXY	Q9WX70 acetobacter
10	114	3.5	692	1 SGA9 HUMAN	P48067 homo sapien
11	113.5	3.5	469	1 NUON BUCAL	P57264 buchnera ap
12	112	3.4	590	1 NU5M TRYBB	P04540 trypanosoma
13	111	3.4	460	1 YAGM ECOLI	P75683 escherichia
14	110.5	3.4	496	1 Y048 DREPA	Q9PR97 ureaplasma
15	110	3.4	633	1 SGA9 MOUSE	P28571 mus musculu
16	109.5	3.3	672	1 AGS8 RAT	P58428 rattus norv
17	109.5	3.3	695	1 FSHR HUMAN	P23945 homo sapien
18	109	3.3	314	1 O1Q1 HUMAN	Q15612 homo sapien
19	108.5	3.3	544	1 ALG6 YEAST	Q12001 saccharomyc
20	108	3.3	408	1 YLO1 SCHPO	Q13883 schizosacch
21	108	3.3	633	1 SGA9 RAT	P28572 rattus norv
22	108	3.3	669	1 NU5M MARPO	P26849 marchantia
23	107.5	3.3	501	1 NUOM BUCAP	Q8K946 buchnera ap
24	107.5	3.3	638	1 SGA9 BOVIN	Q28039 bos taurus
25	107.5	3.3	1465	1 YH85 SCHPO	Q9P590 schizosacch
26	106.5	3.3	426	1 YX35 SCHPO	Q09712 schizosacch
27	106.5	3.3	502	1 PUTP SALTY	P10502 salmonella
28	106	3.2	387	1 SRB3 BRARE	Q91918 brachydanio
29	106	3.2	692	1 NU5C MARPO	P06264 marchantia
30	105.5	3.2	692	1 FSHR RAT	P20395 rattus norv
31	105	3.2	433	1 TCR STAU	P02983 staphylococ
32	104.5	3.2	368	1 CYB TOXGO	Q20672 toxoplasma
33	104.5	3.2	458	1 TCR STRPN	P11063 streptococ

ALIGNMENTS

RESULT 1

ID	YTH3 CAEEL	STANDARD;	PRT;	603 AA.
AC	P54002;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	15-SEP-2003 (Rel. 42, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Putative glycosyl transferase C14A4.3 in chromosome II (EC 2.-.-.-.)			
GN	C14A4.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Pelodierinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	McMurray A.A.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	REVISIONS.			
RA	Durbin R.;			
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (By similarity).			
CC	-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 22.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; Z49909; CAA90107.2; -.			
CC	PIR; T19245; T19245.			
DR	WormPep; C14A4.3; CB32796.			
KW	Hypothetical protein; Transferase; Glycosyltransferase; Transmembrane; Endoplasmic reticulum			
FT	TRANSMEM 109 129			POTENTIAL.
FT	TRANSMEM 131 151			POTENTIAL.
FT	TRANSMEM 167 187			POTENTIAL.
FT	TRANSMEM 196 216			POTENTIAL.
FT	TRANSMEM 246 266			POTENTIAL.
FT	TRANSMEM 311 331			POTENTIAL.
FT	TRANSMEM 344 364			POTENTIAL.
FT	TRANSMEM 371 391			POTENTIAL.
FT	TRANSMEM 398 418			POTENTIAL.
SQ	SEQUENCE 603 AA; 68912 MW; 6AFDD9F995CEC0C5 CRC64;			

Query Match 32.5%; Score 1064; DB 1; Length 603;

Best Local Similarity 39.4%; Pred. No. 6.5e-68;

Matches 226; Conservative 96; Mismatches 212; Indels 40; Gaps 10;

QY 48 NKACQVAPGSGTAFKCLLSARLCALLNISDCDFTFYWEPHYLYYGEFGFTWESP 107

DB 412 LETDNVTT-----NVCVGKWTYRPSFFFLPDNSRLKFKVSEBFDGILPGCFVESNST 464
ZY 508 -----ATRIPTDMNDONLEBPSRYIDISKCHYLVDL--DTMRETPREPKYSSNKEEWS 560
DB 465 WNRREGYQIPENHNEPNEPTYSLSLSCDPLIDLEFDSKATVNEPIYSKS-DGNIP 523
ZY 561 LAYRPFIDARSSKLLRAFYVPLPSQYTVYVNTYILKPRK 601
DB 524 VMVYFFIDTKQTPMGRAFAVPPIEPKMGY-BILVKKPVK 563

RESULT 3

ALG9_YEAST STANDARD; PRT; 555 AA.
AC P53868;
JT 01-OCT-1996 (Rel. 34, Created)
JT 01-OCT-1996 (Rel. 34, Last sequence update)
JT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable mannose 6-phosphate transferase ALG9 (EC 2.4.1.-).
EN ALG9 OR YNL219C OR N1295.
DS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
DX NCBI_TaxID=4932;
XN [1]
XP SEQUENCE FROM N.A.
XC STRAIN=SS328.
XX MEDLINE=96293493; PubMed=8692962;
RA Burda P., Te Heesen S., Brachet A., Wach A., Duesterhoeft A., Abbi M.;
RT "Stepwise assembly of the lipid-linked oligosaccharide in the
RT endoplasmic reticulum of Saccharomyces cerevisiae: identification of
RT the ALG9 gene encoding a putative mannose 6-phosphate transferase";
RL Proc. Natl. Acad. Sci. U.S.A. 93:7160-7165(1996).
XN [2]
XP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE TRANSFER OF MANNOSE FROM DOL-P-MAN TO
CC LIPID-LINKED OLIGOSACCHARIDES.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 22.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X96417; CAA65277.1; -;
DR EMBL; Z71495; CAA96122.1; -;
DR PIR; S63177; S63177.
DR SGD; S0005163; ALG9.
DR GO; GO:0005783; C:Endoplasmic reticulum; IMP.
DR GO; GO:0000030; F:mannosyltransferase activity; IMP.
DR GO; GO:0006486; P:protein amino acid glycosylation; IMP.
KW Transferase; Glycosyltransferase; Transmembrane;
KW Endoplasmic reticulum.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 269 289 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 350 370 POTENTIAL.
FT TRANSMEM 555 AA; 63776 MW; 108ED4E0B0C2AAA7 CRC64;
SQ SEQUENCE

Query Match 22.5%; Score 736; DB 1; Length 555;
Best Local Similarity 33.1%; Pred. No. 9e-45;
Matches 179; Conservative 82; Mismatches 199; Indels 80; Gaps 15;
QY 76 SNLSDCDEFTNWEKPHYLTYGEGFTWEYSPAYAIRSVAYILLHAWPAAFHARLQTNK 135
DB 26 SLISDCDEFTNWEKPHYLTYGEGFTWEYSPAYAIRSVAYILLHAWPAAFHARLQTNK 79
QY 136 IL-----VFYFLRCLLAFVSCICELYFYKAVCKKFGHLSVRMMMLAFVLSTGMFCSSA 189
DB 80 FTDLSEHWNFFITRACLGPFSPFIMEFKLHRETAGSLAQIANIWIIFQLPFGWPHASVE 139
QY 190 FLSSSECMYTTLIAMTGWMDKTSIAVLGVAAG-----AILGWPESSAALG 234
DB 140 LPSAVAMLL-----YVGATRSLRYLSTGSTNSTKSLAYNFLASILGWPPVLILS 191
QY 235 LPIAFDLLVMKRWKSWFFWSLMALILFLV---PYVVIDSYYGKLVIAPLNIVLYNVT 291
DB 192 LPCLHYL-FNHRITITRTAFDCCLIESTLTAFAVIVTDSIFYGKLAPVSNWILFYNVIN 250
QY 292 ---PHGPDLYGTEPWYFYLINGFINVNAFALALLVPLTSLMEVLLQRFHVQNLGHPY 348
DB 251 ASSESGNIFGVEPWYYPNLLNF-----PLPVLVLAIGIFHUR-----LW 294
QY 349 LTLAPYIWIIFIIQPHKEERFLPVVYPLI CLGAVALSALQKCYHFVQRYRLEHYTV 408
DB 295 PLWASLFTWIAVFTQPHKEERFLYIYGLITLSASIAFYKVLNLFN-----RKPI 345
QY 409 TSNWLAGTVFVGLLSFSPRSVALFRGHGDLIDYEPYRIATDTIHTVPEGRVNVCV 468
DB 346 LKXGKILSVLLIVAGAMGRIVALLVNNYTPATVAYEQFSSLNQGGVKAPV-----VNYCT 400
QY 469 GKWEYFPSPSFLPDNWLQFIPSPFRGQLPFPFAE-GPLATRI--VPTDMNDONLEEPS 525
DB 401 GREWYHFPSPSFLPDNHLKFKVSGFDGLPGDFPSSGSIKKIETLPKGMNKNVIYDTG 460
QY 526 RYDIDSKCHYLVDLDMRTPRE---PKYSSNKEEWSLAVRPPFLDARSSKLLRAFYVP 582
DB 461 KEWPIRECYFIDIVAPINLTKDVFNPLHL--MDWNKLAACAAFDGNSKILGRAFYVP 518

RESULT 4

YCO2_YEAST STANDARD; PRT; 616 AA.
ID YG02_YEAST
AC P30777;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 72.6 kDa protein in MRFL-HUL5 intergenic region.
GN YG1142C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1769;
RX MEDLINE=97197983; PubMed=9046099;
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT TIF1, MRFL genes and six new open reading frames.";
RL Yeast 13:177-182(1997).
RN [2]
RP SEQUENCE OF 339-616 FROM N.A.
RX MEDLINE=93117110; PubMed=1475194;
RA Pel H.J., Maat M.J., Rep M., Grivell L.A.;
RT "The yeast nuclear gene MRFL encodes a mitochondrial peptide chain
RT release factor and cures several mitochondrial RNA splicing
RT defects";
RL Nucleic Acids Res. 20:6339-6346(1992).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: SOME, TO S.POMBE SPAC2G11.09, SPAC4G8.12C AND


```
2y 98 EGFOTWEYSPAYAIRSYAYLL-----LHAWPAAFHARILQTNKILVFLRCLLAF 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 KGTIPWEKFAARSGLLVYVGPFTILEFPE-----IQDNPAIILSMRLQNVV 98
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 149 VSCICELYFYKAVCKEGLHVRNMALFLVLS-----TGMPCSSSAFLPSSFCMYTTL 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 NYLLCYTHFLPKLRDRKAVQFQKSLLSLTSVWTYQHTFSNSI-----ETLALISL 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 202 IANTGWTMDKTSIAVLGVAAGAILGWPFSAALGLPIAFDLV-----MKRWKSF-FHNSL 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 TVMEDMNEK-NIORSNFKNSVILGLIFSGVNFVTFPAFIPCLILFWKRYVHWKS 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 257 MALLILFLVP-----VVVIDSYYY-GK-LVIAPLNLVYNVTFPHGPDLYGTETPYFILI 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 FSLLLSFSFSCFLVLDITNINNGKGFVITPLNLKYNL-NVQNLQVHGLHPRYTHLL 272
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 309 NGFLNFVAFALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAPMY-IWFIIFRTQPKH 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 273 -----VNLPOIVGPV--LLAIISGKLDKLS-----TVAILSGLLFISFP--QHO 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 368 EERFLFPVPLICCGAVALSALOKCYHFVQRYLREHYTVTS-NWML----- 415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 ELRLVPLVPL-----VNLNWTPLSLSTLVNKKIEK 346
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 416 GTVFLGGLLSFSRSVALFRGVH---GPLDLYPFPYRIATDPT-IHTVPGRPVNVCKGE 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 GTWLLFNII-----MAFTMGISHQAGIIQIGDYFHFRTQMGVHI----- 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 472 WYRFPSSPLDNLQLOFIPSEFRGQLPKFAEGPLATRIVPTDMNDONLEPS----- 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 WKMTYS-----PPTWNY-----MSNNLTVSSLIINTQDGLSIEDEVAFSGVNH 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2y 526 RYIDISKHYLVLDLWRETPR 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 430 HVIDLKGK-----DLPLLTETIR 447
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
YN81 YEAST
ID YN81 YEAST STANDARD; PRT; 551 AA.
AC P53730;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 46, Last annotation update)
DE Hypothetical 62.7 kDa protein in SEC12-SSK2 intergenic region.
GN YNR030W OR N3265.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO C.ELEGANS ZC513.5 AND S.POMBE SPSC1734.12C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 271645; CAA96310.1; -
CC PIR; S63361; S63361.
CC SGD; S0005313; E0M339.
CC GO; GO:000009; P:alpha-1,6-mannosyltransferase activity; IMP.
CC GO; GO:0006488; P:oligosaccharide-pp-dolichol assembly; IMP.
CC GO; GO:0006486; P:protein amino acid glycosylation; IMP.
CC Hypothetical protein; Transmembrane.
```

```
FT TRANSMEM 3 23
FT TRANSMEM 62 82
FT TRANSMEM 90 110
FT TRANSMEM 137 157
FT TRANSMEM 179 199
FT TRANSMEM 203 223
FT TRANSMEM 228 248
FT TRANSMEM 276 296
FT TRANSMEM 304 324
FT TRANSMEM 332 352
FT TRANSMEM 366 386
FT TRANSMEM 418 438
SQ SEQUENCE 551 AA; 62672 MW; B08A94BBF260502F CRC64;

Query Match 3.9%; Score 128.5; DB 1; Length 551;
Best Local Similarity 21.6%; Pred. NO. 0.076;
Matches 92; Conservative 63; Mismatches 167; Indels 103; Gaps 18;

Qy 60 TAFKCLLSARLCAALLSNISDCDETNWEPHYLYLIGSGFQTEWYSP----- 107
Db 8 TVLLTVISFHLIQAPTKV---EESFNI-QAHDILTVSVFDISQYDHLKFPGVVPRTEV 63
Qy 108 -----AVAIRSYAY---LLHAWPAAFHARIL-----QTNKILVFLRCLLAFVSCIC 153
Db 64 GAVTIAMLSRPYLYLSSLIQTSRPTSIDVQVVRGIVGLTNGLSFIYLNKCLQDMFDEIT 123
Qy 154 E-----LYFYKAVCKEGLHVRNMALFLVLSGTCMFCSSSAFLPSSFCMYTTLIA 203
Db 124 EKKKEENEDKDIYYDSAGTW-----LFLIGSFHLMFYSTRTP-NFVMTLPLTN 174
Qy 204 MT-GW-TYMDKTSIAVLGVAAGAILGWPFSAALGLPIA-FDLLVMKRWKSFPHWSLMALI 260
Db 175 VALGWLLGRYNAAILFLSALVAIVPRLEVSALSGAIFSVIFKKISLFDKIFGIFGLG 234
Qy 261 LFLVPVVVVIDSYVYKGLVITAPLNLVYNVTFPHGPDLYGTETWYF----- 306
Db 235 LGSASITVDSYFWQEWCUPEVDGFLFNVAAGYA-SKWGVEPVYATFYTHLRMMFMPPTV 293
Qy 307 LINGFLNFVAFALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAPMYINFIIFFIQPH 366
Db 294 LLLNYFGYKLAPA-KLKIVSLASL-----FHI-----IVLSFQPH 327
Qy 367 KEERFLFPVPLICCGAVALSALOKCYHFVQRYLREHYTVTSNWLALGTVELGLLSF 426
Db 328 KEWRFIYAYVPSIMLLGATGAHL-----WENMKVKKIT---NVLCLAILPLSLTMSF 377
Qy 427 SRSVA 431
Db 378 FISWA 382

RESULT 7
YADC SCHPO
ID YADC SCHPO STANDARD; PRT; 533 AA.
AC Q09837;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C4G8.12c in chromosome I.
GN SPAC4G8.12C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
```

Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
A Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
A James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
A Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
A Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
A Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
A Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
A Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
A Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
A Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
A Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
A Borzym K., Langer I., Beck A., Lehar H., Reinhardt R., Pohl T.M.,
A Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
A Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,
A Galibert F., Aves S., Xiang Z., Hunt C., Moore K., Hurst S.M.,
A Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
A Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
A Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
A Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
A Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
T The genome sequence of *Schizosaccharomyces pombe*.
L Nature 415:871-880(2002).
C -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
C -1- SIMILARITY: TO YEAST SMP3.
C -1- SIMILARITY: SOME, TO YEAST YGL142C.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
C or send an email to license@isb-sib.ch).
C -----
C EMBL: Z56276; CA91213.1; --
C F.R.; T38857; S62489.
C GeneDB: Srombe; SPAC4G8.12c; --
C InterPro: IPR005599; PMP.
C Pfam: PF03901; PMP; 1.
C Hypothetical protein; Transmembrane.
C TRANSMEM 8 28 POTENTIAL.
T T TRANSMEM 61 81 POTENTIAL.
T T TRANSMEM 91 111 POTENTIAL.
T T TRANSMEM 144 164 POTENTIAL.
T T TRANSMEM 175 195 POTENTIAL.
T T TRANSMEM 216 236 POTENTIAL.
T T TRANSMEM 274 294 POTENTIAL.
T T TRANSMEM 297 317 POTENTIAL.
T T TRANSMEM 338 358 POTENTIAL.
T T TRANSMEM 496 516 POTENTIAL.
T T SEQUENCE 533 AA; 62200 MW; F14519C95884687 CRC64;
Query Match 3.8%; Score 123.5; DB 1; Length 533;
Best Local Similarity 20.1%; Pred. No. 0.16;
Matches 112; Conservative 84; Mismatches 180; Indels 181; Gaps 32;
Y 86 NYNEPHTLYLGSGFO-----TWYSPAYAIRSVAYLLHAWPAAPFARILQ 132
b 26 SYIHPDEHL---OSQIPANKLFGKVKVLPWNETTKPKRS-----VWP 66
Y 133 TNKILVYFPLRCLLAFVSCICELFYKAVCKKFG-LHVSMMMLAFVLVS----- 180
b 67 LNVMLLPFLLC-----RCIC-----KNSCSPYILLFRLYMLCLISLLDLSIWNIVPL 116
Y 181 ----TGNFCSAFLPSF-----CMTTLIAMTGMVMDKTSIAVLG-----V 219
b 117 NAWKALLYSSFMATTFQTHFTNSIETIFFITILFLSKINSVPLNKISLYIFLL 176
Y 220 AAGAILGW-----PFSALGLP-IAPDLYMKHRWS-----FFHSLMALILF-LVPVVV 268
b 177 AIVSVLGFFTRITFLAFVIAPYIVSVRCFKKNNVNPCKDFLHLCIFVSVPATVLACIL 236
Y 269 IDSYIYKLVIAINIVLVNV----FTPGDPLDYGTEPWYFLNGFLNFNVAFALLV 324

237 EDYFYGVFWITWNNKYNQSIENISQHG-----LHSLRTHFTY-----MPLLC 282
QY 325 LPLTSLMEYLQRPFHQVNLGHPYWLTLAPMYWIFIFFIQPKKBERFLPPVVLICLOCA 384
Db 283 GPLI-----FVPLKLDVRK--PATWMLLPVILSLF-----PQEPFLPAAASIF----I 328
QY 385 VALSALOKCY--HFVQRYLEHYVTSTNWLALGTVFLGLLSFS----- 427
Db 329 VNSCLVRSTWIKFLFWY-----AVLAVFGIMHQNGVIAVLEVKNLBQ 376
QY 428 RSAV-----LFRGYHGDLDPYEFVRIADPT-IHTVPEGRPVNVCKGKWRFPSSFL 480
Db 377 RNVTTMENCNLYFNPETTY--FKIYSAPTWMLARPKFSQINT---SHLYNYSYTKQM 431
QY 481 LPDNLWOLF-----IPSEFRGOLPKPFAEGPLATPIV-PTDMNDONLEEPSRYIDISK 532
Db 432 ISKFWETIYEEVKVNVLPPE-----NTTEFOASTLLVCPVAM---LOTSSYLQNLTM 480
QY 533 CHYL---VDLDTMEETP 546
Db 481 LHYIPYHVLDLDDTDELP 497
RESULT 8
CYOB_BUCAI
ID CYOB_BUCAI STANDARD; PRT; 662 AA.
AC P57543;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ubiquinol oxidase polypeptide 1 (EC 1.10.3.-) (Cytochrome O subunit 1)
DE (Oxidase 30(3) subunit 1) (Cytochrome O ubiquinol oxidase subunit 1).
GN CYOB OR BU471.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
synthetic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: CYTOCHROME O TERMINAL OXIDASE COMPLEX IS THE COMPONENT
CC OF THE AEROBIC RESPIRATORY CHAIN THAT PREDOMINATES WHEN CELLS ARE
CC GROWN AT HIGH AERATION. THIS UBIQUINOL OXIDASE SHOWS PROTON PUMP
CC ACTIVITY ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Ubiquinol-8 + O(2) = Ubiquinone-8 + H(2)O.
CC -1- COFACTOR: CONTAINS TWO PROTOHEME IX (HEME B55 AND B562) AND
CC COPPER B (BY SIMILARITY).
CC -1- PATHWAY: UBIQUINOL OXIDASE CATALYZES THE TERMINAL STEP IN THE
CC ELECTRON TRANSPORT CHAIN.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF001119; BAB13168.1; --
CC HSP; P18401; 1FFT.
DR InterPro: IPR000883; COX1.
DR Pfam: PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.

PROSITE, PS00077; COX1; 1.
KW Oxidoreductase; Respiratory chain; Transmembrane; Heme; Copper;
KW Hydrogen ion transport; Complete proteome.
FT DOMAIN 1 14
FT TRANSMEM 15 35
FT TRANSMEM 36 58
FT TRANSMEM 59 79
FT TRANSMEM 80 106
FT TRANSMEM 107 127
FT TRANSMEM 128 145
FT TRANSMEM 146 166
FT TRANSMEM 167 189
FT TRANSMEM 190 210
FT TRANSMEM 211 232
FT TRANSMEM 233 253
FT TRANSMEM 254 277
FT TRANSMEM 278 298
FT TRANSMEM 299 309
FT TRANSMEM 310 330
FT TRANSMEM 331 346
FT TRANSMEM 347 367
FT TRANSMEM 368 380
FT TRANSMEM 381 401
FT TRANSMEM 402 413
FT TRANSMEM 414 434
FT TRANSMEM 435 456
FT TRANSMEM 457 477
FT TRANSMEM 478 493
FT TRANSMEM 494 514
FT TRANSMEM 515 586
FT TRANSMEM 587 607
FT TRANSMEM 608 608
FT TRANSMEM 609 629
FT TRANSMEM 630 662
FT METAL 106 106
FT METAL 284 284
FT METAL 288 288
FT METAL 333 333
FT METAL 334 334
FT METAL 419 419
FT METAL 421 421
FT CROSSLNK 284 288
SQ SEQUENCE 662 AA; 75455 MW; 3F50A2628133EA65 CRC64;
Query Match 3.5%; Score 116; DB 1; Length 662;
Best Local Similarity 21.6%; Pred No. 0.71; Indels 124; Gaps 27;
Matches 108; Conservative 76; Mismatches 191;
70 LCAALLSNISDCDETFNYWEPHYLYGEGFTWEY---SPAYAIRSYAYL-----L 118
29 LCIA-----STITYKKKWLWY-EWFTTVDHKKISIMRGILAFVLMFGRGVDAIL 78
119 LHA-----WPAAFHARILQTNKILVYF--LRCLLAFVSCICELY----- 156
79 METQQVVASAGFKGLPDPHRYDQIFTHAGVIMIFVAVMLVIGLUNLVIPQAGADVAF 138
157 -FYKAVCKKFGHLVSRMMLAFVLSTGMFCSS--AFLPSSFCMYTTLTAMTWYMDKTS 213
139 PELNNL--SFVLNVSSAVLLTSLGIGGFAQTGLAYPLSGIKYSSGVGVYDWIWN--S 193
214 IAVLGVRAAGALILGWFPFSAALGLPTAFDOLLVKKRWKSFH-----WSLMA---LILFLVP 265
194 LQISGVG-----TTLTGINFLVITLKKRAGMGSFFKMPVFTVLSLCTNILLIVISFP 244
266 VVVIDSYYGKLVIAFLNVL--YVFTPHGPDLYGTBPWFYFLINGFLNFVAFALALLV 324
245 VLTV-----TLVLLTDRYENFHFHT---NDLGNAMWYVNLIIWINGHEV---YILV 291
325 LPLTSLMEVLLQRFVQNL-GHP--YMLTAPMWIIFII-----FFTPKHEKRFPPVY 376
292 LPVFGVSEVATFSGKGLFGVSVVATLS:TILSFVWLHFFHTWGADVNTFFGIT 351
377 PLICLCGAVALSALQKCYHFVQVRLEBHYTVTSNWLALGTVFLF---GLLSFSRSVALF 433

Db 352 TMI-----IAIPGVKIPNWLFIYGRVHMISSLILWLGFLVTSIGMTGVLLSVP-- 404
QY 434 RGKHGPLD--LYPEFYRIADPTTIHVPEGRPNVVCVKENYRFPF--SFLLPDNQ--- 486
Db 405 -----PADFVLHNSLFLVA---HFHNVIIIGGVVFCFAGINWFFPKLFGFVLNEIWGKRA 456
QY 487 -----LQFIPSEFRG 496
Db 457 FFWIIGFFLAFLPIYFLG 475
RESULT 9
ID_BCSY_ACEXY STANDARD; PRT; 386 AA.
AC Q9WX70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative membrane-bound transacylase bcsy (EC 2.3.-.-).
GN BCSY.
OS Acetobacter xylinus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Gluconacetobacter.
OX NCBI_TaxID=28448;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 7664 / IFO 13693;
RX MEDLINE=99310341; PubMed=10382968;
RA Umeda Y., Hirano A., Ishibashi M., Akiyama H., Onizuka T., Ikeuchi M., Inoue Y.;
RT "Cloning of cellulose synthase genes from Acetobacter xylinum JCM 7664: implication of a novel set of cellulose synthase genes."
RL DNA Res. 6:109-115(1999).
CC -!- FUNCTION: May acylate a glucose moiety into cellulose fibrils, in cooperation with bcsABII and bcsCII.
CC -!- PATHWAY: Bacterial cellulose biosynthesis.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMEL; AB015803; BAA77595.1; -
CC InterPro; IPR002656; Acyl transf 3.
CC Pfam; PF01757; Acyl transf 3; 1.
KW Cellulose biosynthesis; Transferrase; Acyltransferase; Transmembrane;
KW Inner membrane;
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 91 111 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 290 310 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 362 382 POTENTIAL.
SQ SEQUENCE 386 AA; 42858 MW; AEB53005CEB74B94 CRC64;
Query Match 3.5%; Score 115; DB 1; Length 386;
Best Local Similarity 20.4%; Pred. No. 0.46;
Matches 64; Conservative 50; Mismatches 116; Indels 84; Gaps 15;
QY 50 AGQWABEGSTAFKCLLSARLCAALLSNISDCDETFNYWEPHYLYGEGFTWEYSAY 109
Db 104 AGMLWVQLQSDRADIALNGALVLYLSNI-----WASGH-VGYFQ----- 143

QY 385 VALSALQCYHFVFORVRLHYTTS-NWLALGTGTVFLGGLSFSRVALPRGYHGLDLY 443
Db 622 DGDTLQ-----RLKNATPSRDW-----GPALE 646
QY 444 PFYRIADPTHTVPEGRPNVVCVKWYPPSPFLPDNWQLOFIPSPRQOL 498
Db 647 HRTGRIA--PTIAPSPED-----CFEVSQSLHDKAQPPIVGSNGSRL 687

RESULT 11

NUON BUCAI
ID NUON BUCAI STANDARD; PRT; 469 AA.
AC P57264;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain N (EC 1.6.99.5) (NADH dehydrogenase I, chain N) (NDH-1, chain N).
GN NUON OR BUI166
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-sulfur (Fe-S) centers, to quinones in the respiratory chain.
CC Couples the redox reaction to proton translocation (for every two electrons transferred, four hydrogen ions are translocated across the cytoplasmic membrane), and thus conserves the redox energy in a proton gradient (By similarity).
CC -!- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -!- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOA, H, J, K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AP001118; BAB12884.1; -
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; oxidored_q1; 1.
KW Oxidoreductase; NAD; Quinone; Transmembrane; Complete proteome.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 108 128 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 359 379 POTENTIAL.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 442 462 POTENTIAL.
SQ SEQUENCE 469 AA; 52578 MW; 69370BE48E48C62B CRC64;

Query Match 3.5%; Score 113.5; DB 1; Length 469;
Best Local Similarity 19.7%; Pred. NO. 0.73;

Matches 81; Conservative 76; Mismatches 156; Indels 99; Gaps 18;
QY 126 FHARILQTNKILFYRP-----LRCLAFVSCICELFYKAVCKKFGHVRWMAFLVLS 180
Db 37 FLIIIIINVIIFHNSNAILYIGMILISSICTCTFSPWLLKYFKNKEEYLLVIIST 96
QY 181 TGMFCSSAFLPSPFCMYTTLIAMTWG-----YMDKTS---IAYLGVAAAGAIL-- 225
Db 97 LCAISLTIISHMASFFINIELISLPWFGLIAYSRYQKYSLESSLKYIILSGVSSPFLFG 156
QY 226 -GWPFSAALGLPIAFDLNVKHKWKFHMSLMALILFLVPPVVIDSYYVKLVIAPLNI 284
Db 157 IAWYISISGGL----DFLSI-HKSPNFASEKEILVVLFGISMILLSLF-----KLSI 204
QY 285 VLYNVFTPHGPDLYGTEPW-----YFYLINGFLNF-----NVAF 318
Db 205 VPFLHWT---PDYQGSPTSVLSFFSTAGKISVFSVLLNPLSFSNSDNKIVFILSLII 261
QY 319 ALALLVPLTSLMEYLLOREHVQNLGHPYWLTPAPVWIFLFFIOPHKEERFLPPVPL 378
Db 262 ILSILVGNLALFOKDKRF-----LGY-----TSIQIGVILLVIVSHKNSYFSLEA-SA 312
QY 379 ICLOGAVALSALOKCYHFVFORVRLHYTTSNWLALGTGTVFLGGLSFSRVALPRGYH 438
Db 313 IYLCG-----YLF-----SNIACLGIVNLISTSHINNASSINSYRG 349
QY 439 PLDLYPEFYRIADPTHTVPEGRPNVVCVKWYPPSPFLPDNWQLOF 489
Db 350 LFWSHPLLSSVLTLLVLISSA--GIPMTLGFGR--FYILSVIEHLWLIGF 397

RESULT 12

NUON TRYBB
ID NUON TRYBB STANDARD; PRT; 590 AA.
AC P04540;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
GN ND5.
OS Trypanosoma brucei brucei.
OG Mitochondrion.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85037915; PubMed=6093040;
RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C., van Boom J.H., Benne R.;
RT "The sequence of the gene for cytochrome c oxidase subunit I, a frameshift containing gene for cytochrome c oxidase subunit II and seven unassigned reading frames in Trypanosoma brucei mitochondrial maxi-circle DNA.";
RT Nucleic Acids Res. 12:7327-7344(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=164;
RX MEDLINE=87201680; PubMed=2437452;
RA Jamser D.P., Feagin J.E., Payne M., Stuart K.;
RT "Variation of G-rich mitochondrial transcripts among stocks of Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 22:259-272(1987).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X01094; CAB57807.1; -

```

R EMBL; M14820; AAB59225.1; -.
R PIR; A04519; OQUT5.
R InterPro; IPR003916; NADHox redox.
R InterPro; IPR001750; Oxidized q1.
R Pfam; PF00361; oxidized q1; 1.
R PRINTS; PR01434; NADHGNASE5.
W Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
Q SEQUENCE 590 AA; 71494 MW; CAB231AD9DF5282 CRC64;

Query Match 3.4%; Score 112; DB 1; Length 590;
Best Local Similarity 19.5%; Pred. No. 1.2; Indels 186; Gaps 24;
Matches 94; Conservative 65; Mismatches 138;

Y 114 YAYLLHAWPAFAHRIQTKILVYFRLCLAPV-----SCICEL----- 155
b 93 FCFIVFYAFYMYFDMLLGRFLIIWFIVWCVNLFILSYDLFTAYCGWELLGLFSFLLIS 152
Y 156 -VFYKAVCKKFGHVRWMLAPLVLTSG---MFCSSAPLPSSFCMYT---TLIAMTG 206
b 153 YWYRPFALKFGFK-----AFFIKIGDVLIIAFSIIIFLSNGFCMTTFYFLNFFCMY 206
Y 207 WYMDKTSIAVLGAAGAILGNPFSAAALG-----PIAFLLVMKHR----- 247
b 207 YXTEFSICLLVGA-----FTKSTQFGLHILWLPDAMEGPIPVSAII---HAATIVVCGII 258
Y 248 -----WKSPEH-----WSMALILFVVP-----V 267
b 259 LLSFVWCFDFWFSFYNLGHSTLILMLTLCVYFNEDVKRYVAFSTICQISFSMECL 318
Y 268 VIDSY-----YGLKIVIAPLNIVL---YNVTPHGPDLVYGP- -WYFYL----- 307
b 319 CIDIVIGSLFFCVHMFYKATLFIVLGIWHIF-----FGLQDLRCVFFMYFGCVLAR 371
Y 308 -----INGFLNVAFAALIVLPLTSLMEVLLQRFHVNQILGHYPWLTLA 352
b 372 LLLIFAILNCSWFLCGLFYCKGMILLAL-LMLLSYNIIEFLISI-----IFIFT 422
Y 353 PMYINFIIFIQPKHEERLEFPVPLI-CL-----CGAVALSALQKV----- 394
b 423 MIYNYELLFFLM-----FVFKCFCLVDCFLFLFDYECVCLVCLISLYMCLTSIFFIIDF 476
Y 395 -----HFVQRYLEHYTVTSNWLALGTFLPGLLSFSRSVALRGVHGPDLDLYPEF 446
b 477 VCIFFPSSYCVFWSFFLNFYNFDF--IAIFVFLILSVGLVGYGLCF-FYFFNIDCIMLF 533
Y 447 YRI 449
b 534 WRI 536

RESULT 13
YAGG ECOLI
D YAGG ECOLI STANDARD; PRT; 460 AA.
C P75683; P71292;
T 01-NOV-1997 (Rel. 35, Created)
T 01-NOV-1997 (Rel. 35, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
T Hypothetical symporter yagg.
N N YAGG OR B0270.
S Escherichia coli.
S Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
S Enterobacteriaceae; Escherichia.
X NCBI_TaxID=562;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=K12 / MG1655;
X MEDLINE=97426617; PubMed=9278503;
A Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
A Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
A Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
A Mau B., Shao Y.;
A "The complete genome sequence of Escherichia coli K-12."
A Science 277:1453-1474(1997).

```

```

[2]
RN SEQUENCE FROM N.A.
RP Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC (SGP). STRONG, TO E.COLI YICJ.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE000135; AAC73373.1; -.
DR EMBL; U70214; AAB08691.1; -.
DR PIR; F64752; F64752.
DR EcoGene; EG13346; yagg.
DR InterPro; IPR001927; Na/Gal symport.
DR TIGRFAMs; TIGR00792; gph; 1.
DR PROSITE; PS00872; NA_GALACTOSIDE_SYMP; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Symport; Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 107 127 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 309 329 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 370 370 K -> N (IN REF. 2).
FT CONFLICT 370 370
SQ SEQUENCE 460 AA; 50631 MW; AB600589E663C6B7 CRC64;

Query Match 3.4%; Score 111; DB 1; Length 460;
Best Local Similarity 22.7%; Pred. No. 1.1; Indels 82; Gaps 16;
Matches 67; Conservative 46; Mismatches 100;

QY 195 FCMYTTLIAMTGWYMDKTSIAVLGAAGAILGNPFSAAALGPIAFD---LLYMKHR--- 247
Db 21 FVWQATMELLAYFVTD-----VFGLSAG-IMGTLFLVSRVLDVATDPLMGLLVDRTRH 74
QY 248 --WKSPEH-----SLMALILFVVPVVVIDSYYYKLVIAPLNIVL----- 287
Db 75 GQFRPFLWGAIPFGIVCVLTFYTP---DFSAGKIIYACVTVILTLYTVTFVNVPYCA 130
QY 288 --NVFTPHGPDLVYGPWYFYFLINGFLNFNVAFAAL--LVPLTSLM----EYLLQRFH 339
Db 131 MFGVITADPKERHALQSWRFELA-----AAGSLAISGIALPLVSIIGKGE-----Q 177
QY 340 VQNLGHPWTLIAPMYIWFIIFFIOPKHEERLEFPVVP-----LILCG-----AVA 386
Db 178 VGYFGAMCVLGLSGVILLYVCFET---TKERYTEFVQPGSSVAKDLKLLGNSQWRIMCA 234
QY 387 LSAKQKCYHFV-----FQYRLHEHYTVTSNWLALGTV-FLFGLLSFSRSVALF 433
Db 235 FPMATCSNVVRGGATLVFYKYVMDHPELATQFLLYGSLATMFGSLCSSLGLGHF 289

RESULT 14
Y048 UREPA
ID Y048 UREPA STANDARD; PRT; 496 AA.
AC Q9PR37;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

```
16-OCT-2001 (Rel. 40, Last annotation update)
Hypothetical protein UU048.
Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
[1]
SEQUENCE FROM N.A.
STRAIN=Serovar 3;
MEDLINE=30500219; PubMed=11048724;
Glass J.I.; Lefkowitz B.J.; Glass J.S.; Heiner C.R.; Chen E.Y.;
Cassell G.H.;
"The complete sequence of the mucosal pathogen Ureaplasma
urealyticum."
Nature 407:757-762(2000).
- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; AE002104; AAF30453.1;
Hypothetical protein; Transmembrane; Complete proteome.
TRANSMEM 33 53 POTENTIAL.
TRANSMEM 89 109 POTENTIAL.
TRANSMEM 127 147 POTENTIAL.
TRANSMEM 154 174 POTENTIAL.
TRANSMEM 193 213 POTENTIAL.
TRANSMEM 247 267 POTENTIAL.
TRANSMEM 285 305 POTENTIAL.
TRANSMEM 320 340 POTENTIAL.
TRANSMEM 355 375 POTENTIAL.
TRANSMEM 382 402 POTENTIAL.
TRANSMEM 411 431 POTENTIAL.
TRANSMEM 455 475 POTENTIAL.
SEQUENCE 496 AA; 59349 MW; 4F2330427DC3525D CRC64;
Query Match 3.4%; Score 110.5; DB 1; Length 496;
Best Local Similarity 19.8%; Pred. No. 1.3;
Matches 78; Conservative 63; Mismatches 131; Indels 121; Gaps 17;
95 IYGGFQWEYSPAYATRSAYVLLHAWPAAFHARILQTNKILVFVLEKCLAF----- 148
57 LFLGFGDENKSMISKLNAPFDLHS-----PKYQLNPLVIFRFP--LSFPLFVTL 108
149 VSCICELYFYKAVCKFGLHVSRLAFVLSTGMPCSSAFPLPSFCMYTTLIAMTQWY 208
109 IKNTNLYWHRATIKKY---LPWFVLYLVIA-----ISFLFFTFPSV---- 149
209 MDKTSIAVLGVAGAILGWPFSAALGLPIAFDPLLVMKRWKSFHWSLWALLFLVPPV-V 267
150 -----WP-----KEVNLVFLLLVIFLNLISY 171
268 VIDSYY-----YGKLVIAPL---NIVLYNVFTPHGPDLYGTEFPWFYFLINGFL 312
172 EIFNYFISKTNPLLYDNYKNLIAMVFOALLLFLVITPLVWINTGKSPNLFVDNRFY 231
313 -----NFNVA-----FALALLVPLTSLMEVLLQRFVQV--LGHFYV---LT 350
232 TRIVDITVOSGKRNFIILIAFFFLITFIIVLANTNPFALVINKRYDRYVKNLFIILL 291
351 LAPMYIWFIFPIOPHKEERF-----LFPVYPLICLCAVAL-----SALQKCYHFVFQ 399
292 PSAIFMLRLRVFAKHNENLPVGNHLLWVYLQSFPAIILILVWVTLKRLSVKSS 351
400 RYRLEHTVTSNWLALGTVFLFGLLSFRSVAL 432
352 LNTLLNLVVTQTLSL-SLFLVTLFNSKSVSL 383
```

RESULT 15
S6A9_MOUSE
ID S6A9_MOUSE STANDARD; PRT; 633 AA.
AC P28571;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sodium- and chloride-dependent glycine transporter 1 (GlyT1) (GlyT-1).
GN SLC6A9 OR GLYT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS GLYT-1A AND GLYT-1B).
RC TISSUE=Brain;
RX MEDLINE=92316254; PubMed=1618338;
RA Liu Q.-R.; Nelson H.; Mandlyan S.; Lopez-Corcuera B.; Nelson N.;
RT "Cloning and expression of a glycine transporter from mouse brain."
RL FBS Lett. 305:110-114(1992).
CC -!- FUNCTION: Terminates the action of glycine by its high affinity
CC sodium-dependent reuptake into presynaptic terminals. May play a
CC role in regulation of glycine levels in NMDA receptor-mediated
CC neurotransmission.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=GlyT-1A;
CC IsoId=P28571-1; Sequence=Displayed;
CC Name=GlyT-1B;
CC IsoId=P28571-2; Sequence=VSP_006272;
CC -!- SIMILARITY: Belongs to the sodium:neurotransmitter symporter (SNP)
CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X67056; CAA47440.1; -
PIR; S23151;
MD; MG1:95760; SLC6A9.
InterPro; IPR000175; Na/ntran_symport.
Pfam; PF00209; SNF; 1.
PRINTS; PR00176; NANEUSMPORT.
ProDom; PD00448; Na/ntran_symport; 2.
ProSite; PS00610; NA_NEUOTRAN_SYMP_1; 1.
ProSite; PS00754; NA_NEUOTRAN_SYMP_2; 1.
ProSite; PS00754; NA_NEUOTRAN_SYMP_3; 1.
ProSite; PS00267; NA_NEUOTRAN_SYMP_4; 1.
Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
Symport; Amino-acid transport; Alternative splicing.
DOMAIN 1 35 CYTOPLASMIC (POTENTIAL).
TRANSMEM 36 56 1 (POTENTIAL).
TRANSMEM 64 83 2 (POTENTIAL).
TRANSMEM 107 127 3 (POTENTIAL).
DOMAIN 128 214 EXTRACELLULAR (POTENTIAL).
TRANSMEM 215 233 4 (POTENTIAL).
TRANSMEM 242 259 5 (POTENTIAL).
TRANSMEM 295 312 6 (POTENTIAL).
TRANSMEM 324 345 7 (POTENTIAL).
TRANSMEM 378 397 8 (POTENTIAL).
TRANSMEM 426 444 9 (POTENTIAL).
TRANSMEM 460 480 10 (POTENTIAL).
TRANSMEM 501 520 11 (POTENTIAL).
TRANSMEM 539 557 12 (POTENTIAL).
DOMAIN 558 633 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
1 CARBOHYD 183 183 N-LINKED (GLCNAC... ) (POTENTIAL).
2 VARSPLIC 1 10 MVGGARGML -> MASAQGVATPSFEQ (in isoform
3 GLYT-1B).
4 /FTid=VSP_006272.
5 SEQUENCE 633 AA; 70775 MW; FD64AA561F470793 CRC64;
Query Match 3.4%; Score 110; DB 1; Length 633;
Best Local Similarity 20.9%; Pred. No. 1.8;
Matches 77; Conservative 43; Mismatches 144; Indels 104; Gaps 15;
84 TPNWEPHYLIYG---EGFQWEYSPAYAIRSYAYILLHANPAAFHARILQTNKILVPY 140
248 TFPYVVLITLFRGVTLLEG-----AFTGIMYYLTPQWDKYLEAKVWGDAASQIFY 297
141 PLRC-----LLAFVSCICELY-----FYKAVCKKFGGLHVS 170
298 SLGCWGLITWASYNKHNNCYRDSVITNCATRLYAGFVIFSLGFMANHLGVDVS 357
171 RWM-----LAFVLVLTGM-----FCSSSAFLPSSFCMYTTLIAM-----T 205
358 RVADHGPGGLAFVAYPEALTLPISPLWSLLFFFMILLGLGTQFCLETLIVTAIVDEGN 417
206 GYMDKTSIATLVGA-AGAILGWPFSAALGLPIAFDILLVNMKHWKSPFFHWSLMALILELV 264
418 EWILQKTYVTILGVAVAGFLGILTSQAGI---YWLLMDNTAASF-----SLVVIS 468
265 PVWVIDSYYYGKLVIAPLNIVLVNFTPHGPDLYGTEPMYFYILINGPLNPNVAFALALIV 324
469 INCVSIWYIGH-----RNVFQDIQWMLGFPPLFFQICWRF---VSPAIIFFI 514
325 LPLTSLMEYLLQRFHVONLGHVWLTAPMYIWIIFFIQPHKEERPLFPVPELICLCGA 384
515 LIFTVIOYRPITYNHQ---YPGW-----AVRIGFIMAL-----SSVICIPLVAFQLCRT 562
385 VALSALQK 392
563 DGDITLQR 570
```

Search completed: January 13, 2004, 07:02:08
db time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 13, 2004, 06:32:04 ; Search time 59 Seconds

(without alignments)

2672.375 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGASSGDT.....VNTILKPRKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvrius:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3274	100.0	611	4 Q9H068	Q9h068 homo sapien
2	3268	99.8	611	4 Q96P93	Q96pb9 homo sapien
3	3260.5	99.6	618	4 Q96G57	Q96g87 homo sapien
4	3259.5	99.6	618	4 Q9H6U8	Q9h6u8 homo sapien
5	3043	92.9	611	11 Q8VDI9	Q8vdi9 mus musculus
6	3039	92.8	611	11 Q8C7G0	Q8c7g0 mus musculus
7	2739	83.7	570	11 Q8C378	Q8c378 mus musculus
8	1358	41.5	623	5 Q9VBV8	Q9vbv8 drosophila
9	997	30.5	198	6 Q9BGV0	Q9bgv0 macaca fasc
10	849.5	25.9	596	10 Q8LRH5	Q8lrh5 oryza sativ
11	812	24.8	570	10 Q9FZ49	Q9fz49 arabidopsis
12	376	11.5	78	11 Q8BT44	Q8bt44 mus musculus
13	296.5	9.1	548	10 Q9A15	Q9a15 arabidopsis
14	278.5	8.5	498	10 Q9LEQ5	Q9leg5 arabidopsis
15	264.5	8.1	496	5 Q967F2	Q967f2 caenorhabdi
16	235	7.2	561	5 Q9VZM5	Q9vzm5 drosophila

17	234	7.1	506	10 Q9AWW0	Q9aww0 oryza sativ
18	219.5	6.7	558	5 Q9NKKZ7	Q9nkz7 trypanosoma
19	209.5	6.4	554	4 Q92521	Q92521 homo sapien
20	208.5	6.4	554	4 Q8WVN7	Q8wvn7 homo sapien
21	207	6.3	506	3 Q9USN0	Q9usn0 schizosacch
22	194.5	5.9	542	11 Q9JQ0	Q9jq0 mus musculus
23	193	5.9	492	5 Q23361	Q23361 caenorhabdi
24	183	5.6	509	2 Q9RNP1	Q9rnp1 zymomonas m
25	182.5	5.6	488	4 Q96AA4	Q96aa4 homo sapien
26	181.5	5.5	488	4 Q9BV10	Q9bv10 homo sapien
27	181.5	5.5	683	4 Q8NG10	Q8ng10 homo sapien
28	180.5	5.5	678	5 Q9VH78	Q9vh78 drosophila
29	168	5.1	696	5 Q8MT80	Q8mt80 drosophila
30	168	5.1	830	5 Q9W176	Q9w176 drosophila
31	161	4.9	483	11 Q9VDB2	Q9vdb2 mus musculus
32	141	4.3	547	3 Q74753	Q74753 schizosacch
33	135.5	4.1	603	2 Q8KJD8	Q8kjd8 rhizobium l
34	132.5	4.0	491	2 Q9EV68	Q9ev68 rhizobium m
35	131.5	4.0	487	16 Q9JS51	Q9js51 chlamydia p
36	131.5	4.0	487	16 Q929E6	Q929e6 chlamydia p
37	131.5	4.0	491	16 Q92YNE	Q92yne rhizobium m
38	130.5	4.0	583	16 Q9FEU9	Q9feu9 xylella fas
39	130	4.0	433	16 P1970	P1970 mycobacteri
40	129	3.9	477	16 Q8YB02	Q8yb02 brucella me
41	127.5	3.9	421	2 Q9L9P3	Q9l9p3 lactococcus
42	127	3.9	465	17 Q50564	Q50564 methanobact
43	127	3.9	537	16 Q98LJ8	Q98lj8 rhizobium l
44	127	3.9	718	17 Q97ZT2	Q97zt2 sulfobolus
45	126.5	3.9	1962	5 Q46669	Q46669 canis famil

ALIGNMENTS

RESULT 1

Q9H068	PRELIMINARY;	PRT;	611 AA.
ID Q9H068			
AC Q9H068;			
DT 01-MAR-2001 (TREMBLrel. 16, Created)			
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE Hypothetical protein.			
GN DKFZP586M2420.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Uterus;			
RX MEDLINE=21154917; PubMed=11230166;			
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,			
RA Ansorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,			
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,			
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,			
RA Wamburt R., Korn B., Klein M., Poustka A.;			
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and			
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."			
RL Genome Res. 11:422-435(2001).			
DR EMBL; AL136927; CAB65861.1; -			
KW Hypothetical protein.			
SQ SEQUENCE 611 AA; 69863 MW; 51EC72DDBD866713 CRC64;			

Query Match	100.0%;	Score 3274;	DB 4;	Length 611;
Best Local Similarity	100.0%;	Pred. No. 3.7e-257;		
Matches 611;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1 MASRGARQLKSGASSGDTAPAADKRLRLSGRAGGAHRTLSGNKAGQVWAPGSGT 60			
Db	1 MASRGARQLKSGASSGDTAPAADKRLRLSGRAGGAHRTLSGNKAGQVWAPGSGT 60			
QY	61 AFKCLLSARLCALLSNISDCDETNWPEPTHVLYIGEGFTWEXSPAYIRSYAVLLIH 120			

```
61 AFKLLSARLCAALLSNISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIRSVAYLLH 120
121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
181 TGMFCSSAFILPSSFCMYTTLIANTGMYNDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
181 TGMFCSSAFILPSSFCMYTTLIANTGMYNDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
241 LLVMKRWKSFHWSLMALILFLVPVVVWIDSYKGLVIAPLNVLVYNVFTPHGPDLYGT 300
241 LLVMKRWKSFHWSLMALILFLVPVVVWIDSYKGLVIAPLNVLVYNVFTPHGPDLYGT 300
301 EPWYFYLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII 360
301 EPWYFYLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII 360
361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCVHFVQRYRLEHYTVTSNMLALGTVEL 420
361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCVHFVQRYRLEHYTVTSNMLALGTVEL 420
421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
481 LPDNWLOQFIPEFRGOLPKPFAAGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
481 LPDNWLOQFIPEFRGOLPKPFAAGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
541 TMRETPREPKYSSNKEEWISLAYRPFLDASRSKLLRAFYVPFLSDQYTVVYNYTILKPR 600
541 TMRETPREPKYSSNKEEWISLAYRPFLDASRSKLLRAFYVPFLSDQYTVVYNYTILKPR 600
601 KAKQIRKXSGG 611
601 KAKQIRKXSGG 611
```

RESULT 2

```
96PB9 Q96PB9 PRELIMINARY; PRT; 611 AA.
AC Q96PB9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DIBD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
KW DIBD1, a novel gene homologous to Saccharomyces cerevisiae ALG9, is
   disrupted in a family with bipolar affective disorder by a
   translocation breakpoint at 11q23.;
L EMBL; AF39532; AAL25798.1; -.
Q SEQUENCE 611 AA; 69877 MW; 2736B6750B5ED302 CRC64;
```

Query Match 99.8%; Score 3268; DB 4; Length 611;
Best Local Similarity 99.8%; Pred. No. 1.1e-256;
Matches 610; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
1 MASRGARQRLKSGSGSGDTAPAAKRLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
1 MASRGARQRLKSGSGSGDTAPAAKRLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
61 AFKLLSARLCAALLSNISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIRSVAYLLH 120
```

```
Db 61 AFKLLSARLCAALLSNISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIRSVAYLLH 120
Qy 121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
Db 121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
Qy 181 TGMFCSSAFILPSSFCMYTTLIANTGMYNDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
Db 181 TGMFCSSAFILPSSFCMYTTLIANTGMYNDKTSIAVLGVAAGAILGWPFSALGLPIAFD 240
Qy 241 LLVMKRWKSFHWSLMALILFLVPVVVWIDSYKGLVIAPLNVLVYNVFTPHGPDLYGT 300
Db 241 LLVMKRWKSFHWSLMALILFLVPVVVWIDSYKGLVIAPLNVLVYNVFTPHGPDLYGT 300
Qy 301 EPWYFYLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII 360
Db 301 EPWYFYLINGFLNFVAPALALLVPLTSLMEYLLQRFHVQNLGHPYWLTLAPMYIWFII 360
Qy 361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCVHFVQRYRLEHYTVTSNMLALGTVEL 420
Db 361 FFIQPHKEERFLFPVYPLICLGAVALSALQKCVHFVQRYRLEHYTVTSNMLALGTVEL 420
Qy 421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
Db 421 FGLLSFSRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPNVVCVKWYRFPSSFL 480
Qy 481 LPDNWLOQFIPEFRGOLPKPFAAGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
Db 481 LPDNWLOQFIPEFRGOLPKPFAAGPLATRIVPTDMNDQNLSEPSRYIDISKCHYLVDLD 540
Qy 541 TMRETPREPKYSSNKEEWISLAYRPFLDASRSKLLRAFYVPFLSDQYTVVYNYTILKPR 600
Db 541 TMRETPREPKYSSNKEEWISLAYRPFLDASRSKLLRAFYVPFLSDQYTVVYNYTILKPR 600
Qy 601 KAKQIRKXSGG 611
Db 601 KAKQIRKXSGG 611
```

RESULT 3

```
Q96GS7 Q96GS7 PRELIMINARY; PRT; 618 AA.
AC Q96GS7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
KW EMBL; BC009255; AAH09255.1; -.
DR Hypothetical protein.
SQ SEQUENCE 618 AA; 70786 MW; 05790C3513BF8A4D CRC64;
```

Query Match 99.6%; Score 3260.5; DB 4; Length 618;
Best Local Similarity 98.9%; Pred. No. 4.7e-256;
Matches 611; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

```
Qy 1 MASRGARQRLKSGSGSGDTAPAAKRLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
Db 1 MASRGARQRLKSGSGSGDTAPAAKRLRELLGSRAGGAHRTLSGNKAGQWAPGSGT 60
Qy 61 AFKLLSARLCAALLSNISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIRSVAYLLH 120
Db 61 AFKLLSARLCAALLSNISDCDETFNWEPTHYLYYGEQFTWEYSPAYAIRSVAYLLH 120
Qy 121 AWPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGHLVSRMMLAFVL 180
```


58 GSTAFKCLLSARLCAALLSNISDCDETFNTWEPHTHYLYVGEQOTWEYSPAYAIRSYAYL 117
 58 GSTAFKCLLSARLCAALLSNISDCDETFNTWEPHTHYLYVGEQOTWEYSPAYAIRSYAYL 117
 118 LHWAPAAFAHRIIQTQNKILVFLRCLLAFVSCICELFYKAVCKKFGHVSRRMLAFL 177
 118 LHWAPAAFAHRIIQTQNKILVFLRCLLAFVSCICELFYKAVCKKFGHVSRRMLAFL 177
 178 VLSGTMFCSSAFIPSSFCMYTTIANTGMYMDKTSIAVLGVAAGAILGWPFSAALGLPI 237
 178 VLSGTMFCSSAFIPSSFCMYTTIANTGMYMDKTSIAVLGVAAGAILGWPFSAALGLPI 237
 238 AFDLLVMKRWKSFHNSLMALILFLVPVVVDSYYGKLVIAPLNVLVNYFTSHGPD 297
 238 AFDLLVMKRWKSFHNSLMALILFLVPVVVDSYYGKLVIAPLNVLVNYFTSHGPD 297
 298 YGTEPWYFYLYNGFLNFVAFALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLAPMYIW 357
 298 YGTEPWYFYLYNGFLNFVAFALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLAPMYIW 357
 358 FIIFIQPHKEERFLFPVPLICLCAVALSALOKCYHFVQRYLREHYTVTSNWLALGT 417
 358 FIIFIQPHKEERFLFPVPLICLCAVALSALOKCYHFVQRYLREHYTVTSNWLALGT 417
 418 VFLGLLSFRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVKWYRPPS 477
 418 VFLGLLSFRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVKWYRPPS 477
 478 SFLLPDNWLQFIPSEFRGOLPKPAEGPLATRIPTVPTMDNDQNLSEPSRYIDISKCHYL 537
 478 SFLLPDNWLQFIPSEFRGOLPKPAEGPLATRIPTVPTMDNDQNLSEPSRYIDISKCHYL 537
 538 DLDTWRETPREPKYSSNKEEWSIAYRPFDASSKLLRAFYVPFLSDQYTVVNYTIL 597
 538 DLDTWRETPREPKYSSNKEEWSIAYRPFDASSKLLRAFYVPFLSDQYTVVNYTIL 597
 598 KPRKAKOIRKXSGG 611
 598 KPRKAKOIRKXSGG 611

RESULT 6
 Q8C7G0 PRELIMINARY; PRT; 611 AA.
 ID Q8C7G0
 AC Q8C7G0
 DT 01-MAR-2003 (TremBrel. 23, Created)
 DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
 DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
 DE DIBD1 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 S SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=liver;
 MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 A the RIKEN Genome Exploration Research Group Phase I & II Team;
 T "Analysis of the mouse transcriptome based on functional annotation of
 T 60,770 full-length cDNAs."
 L Nature 420:563-573 (2002).
 R EMBL; AK050335; BAC34195.1;
 Q SEQUENCE 611 AA; 69588 MW; 6298C0C6D0BB1742 CRC64;

Query Match 92.8%; Score 3039; DB 11; Length 611;
 Best Local Similarity 93.3%; Pred. No. 4.3e-238;
 Matches 573; Conservative 10; Mismatches 25; Indels 6; Gaps 4;
 Y 1 MASRRARORLKL---GSGASSGDTAPADKRLLELGSREAGGAEHTELSGNKAGQVWAPE 57
 b 1 MASRRARORLKLKGSGGGGGGGGAGPAAEKL-EQLGSREA-GAEPRPE-SGNKAGQVWAPE 57

58 GSTAFKCLLSARLCAALLSNISDCDETFNTWEPHTHYLYVGEQOTWEYSPAYAIRSYAYL 117
 58 GSTAFKCLLSARLCAALLSNISDCDETFNTWEPHTHYLYVGEQOTWEYSPAYAIRSYAYL 117
 118 LHWAPAAFAHRIIQTQNKILVFLRCLLAFVSCICELFYKAVCKKFGHVSRRMLAFL 177
 118 LHWAPAAFAHRIIQTQNKILVFLRCLLAFVSCICELFYKAVCKKFGHVSRRMLAFL 177
 178 VLSGTMFCSSAFIPSSFCMYTTIANTGMYMDKTSIAVLGVAAGAILGWPFSAALGLPI 237
 178 VLSGTMFCSSAFIPSSFCMYTTIANTGMYMDKTSIAVLGVAAGAILGWPFSAALGLPI 237
 238 AFDLLVMKRWKSFHNSLMALILFLVPVVVDSYYGKLVIAPLNVLVNYFTSHGPD 297
 238 AFDLLVMKRWKSFHNSLMALILFLVPVVVDSYYGKLVIAPLNVLVNYFTSHGPD 297
 298 YGTEPWYFYLYNGFLNFVAFALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLAPMYIW 357
 298 YGTEPWYFYLYNGFLNFVAFALALLVPLTSLMEYLLQRFHVQNLGHPYMLTLAPMYIW 357
 358 FIIFIQPHKEERFLFPVPLICLCAVALSALOKCYHFVQRYLREHYTVTSNWLALGT 417
 358 FIIFIQPHKEERFLFPVPLICLCAVALSALOKCYHFVQRYLREHYTVTSNWLALGT 417
 418 VFLGLLSFRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVKWYRPPS 477
 418 VFLGLLSFRSVALFRGYHGPLDLYPEFYRIATDPTIHTVPEGRPVNVCVKWYRPPS 477
 478 SFLLPDNWLQFIPSEFRGOLPKPAEGPLATRIPTVPTMDNDQNLSEPSRYIDISKCHYL 537
 478 SFLLPDNWLQFIPSEFRGOLPKPAEGPLATRIPTVPTMDNDQNLSEPSRYIDISKCHYL 537
 538 DLDTWRETPREPKYSSNKEEWSIAYRPFDASSKLLRAFYVPFLSDQYTVVNYTIL 597
 538 DLDTWRETPREPKYSSNKEEWSIAYRPFDASSKLLRAFYVPFLSDQYTVVNYTIL 597
 598 KPRKAKOIRKXSGG 611
 598 KPRKAKOIRKXSGG 611

RESULT 7
 Q8C378 PRELIMINARY; PRT; 570 AA.
 ID Q8C378
 AC Q8C378
 DT 01-MAR-2003 (TremBrel. 23, Created)
 DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
 DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
 DE DIBD1 homolog.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 S SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Head;
 MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 A the RIKEN Genome Exploration Research Group Phase I & II Team;
 T "Analysis of the mouse transcriptome based on functional annotation of
 T 60,770 full-length cDNAs."
 L Nature 420:563-573 (2002).
 R EMBL; AK086674; BAC39717.1;
 Q SEQUENCE 570 AA; 64393 MW; 8FFAE5572461A117 CRC64;

Query Match 83.7%; Score 2739; DB 11; Length 570;
 Best Local Similarity 93.5%; Pred. No. 8.6e-214;
 Matches 517; Conservative 5; Mismatches 25; Indels 6; Gaps 4;
 Y 1 MASRRARORLKL---GSGASSGDTAPADKRLLELGSREAGGAEHTELSGNKAGQVWAPE 57
 b 1 MASRRARORLKLKGSGGGGGGGGAGPAAEKL-EQLGSREA-GAEPRPE-SGNKAGQVWAPE 57

```
QY 58 GSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYIGEGFQWTSYPAIRSVAYL 117
DB 58 GSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYIGEGFQWTSYPAIRSVAYL 117
QY 118 LHAWPAFAHARILQTNKILVFLRCLLAFVSCICELYPFKVACKFGLHVRMMLAF 177
DB 118 LHAWPAFAHARILQTNKILVFLRCLLAFVSCICELYPFKVACKFGLHVRMMLAF 177
QY 178 VLSGTGFCSSSAFLPSSFCMTYLLIAMTGMWMDKTSIAVLGVAAGAILGWPFAALGLPI 237
DB 178 VLSGTGFCSSSAFLPSSFCMTYLLIAMTGMWMDKTSIAVLGVAAGAILGWPFAALGLPI 237
QY 238 AFDLVLMKRWKSPFHSNLMALILFLVPVVDISYYTKLVIAPLNVLNVNTPHGPDL 297
DB 238 AFDLVLMKRWKSPFHSNLMALILFLVPVVDISYYTKLVIAPLNVLNVNTPHGPDL 297
QY 298 YCTEPWYFLNGFLNFVAFALALLVPLTSLMEYLQRFVONLGHYPWYLTAPWYIW 357
DB 298 YCTEPWYFLNGFLNFVAFALALLVPLTSLMEYLQRFVONLGHYPWYLTAPWYIW 357
QY 358 FIIFIQPKERFLFPVYPLICLGAVALSALOKCHYFQYRLRLEHYVTWGNALGT 417
DB 358 FIIFIQPKERFLFPVYPLICLGAVALSALOKCHYFQYRLRLEHYVTWGNALGT 417
QY 418 VFLFGLSFSRVALFRGVHGPLDLPEFYRIADPTIHTVPEGRPVNVCWKWYRFP 477
DB 418 VFLFGLSFSRVALFRGVHGPLDLPEFYRIADPTIHTVPEGRPVNVCWKWYRFP 477
QY 478 SFLLPDNMQLOIPSEFRGOLPKPPAEGPLATRIPTDMNDQNLPEPSRYIDISKCHYL 537
DB 478 SFLLPDNMQLOIPSEFRGOLPKPPAEGPLATRIPTDMNDQNLPEPSRYIDISKCHYL 537
QY 538 DLDWTRETPREPK 550
DB 538 DLDWTRETPSGAK 550
```

RESULT 8

```
Q9VBV8 PRELIMINARY; PRT; 623 AA.
ID Q9VBV8
AC Q9VBV8
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 20, Last annotation update)
DE CG11851 protein (LD40966p).
GN CG11851.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.N., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
```

```
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2183-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Pargasa V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003751; AAF56419.1; -.
DR EMBL; AY069643; AAL39788.1; -.
DR FlyBase; FBgn0039293; CG11851.
SQ SEQUENCE 623 AA; 72009 MW; 97FA5120BD432850 CRC64;
```

Query Match 41.5%; Score 1358; DB 5; Length 623;

Best Local Similarity 42.4%; Pred. No. 1e-101;

Matches 267; Conservative 115; Mismatches 213; Indels 34; Gaps 10;

```
QY 1 MASRGARQLKSGSSGSDTAPAADKLRELLG---SREAGGAHRTELSTGNKAGQWA-- 55
DB 1 MAPFAARARVIANKADNQILSKPPKPPGLNGNNKTKATPAGKKDKDAKRNQTSGG 60
QY 56 -----PEGSTAFKCLLSARLCAALLSNISDCDETFNWEPTHYLYIGEGFQWTSY 105
DB 61 QEKGLPNFIMPSTQPTAKPTVTSARLCSAIWAYIADCDTFNWEPTHYLYINGHGLTWEY 120
QY 106 SPAYAIRSYAVLLHAWPAFAHARILQTNKILVFLRCLLAFVSCICELYPFKVACKF 165
DB 121 SPQGLRSYTYLLQGVGFYOKLFNPSPIILFYVRCMLGFGCAWMERYMYKSCQEF 180
QY 166 GLHVSRLMLAFVLSTGMFCSSSAFLPSSFCMTYLLIAMTGMWMDKTSIAVLGVAAGAIL 225
DB 181 GIHIGRLWLI.FQLPVGVMFVSSTALLPSSFSMYFGCAALAAWQQNYCFALFILTALSALL 240
QY 226 GWPSAALGLPIAFDLVLMKRWKSPFHSNLMALILFLVPVVDISYYTKLVIAPLNVL 285
DB 241 GWPFAALIGLPLVLEMLLRDWDKTFVQMLTSGATVAPMTALDTSYFGKLTFAPLNVL 300
QY 286 LYNVFTPHGDLXGTEPWYFLNGFLNFVAFALALLVPLTSLMEYLQRFVONLGHYPWY 344
DB 301 WYNVFTSHGNIFCTEPLSYIINGFLNFNIWLLA-LQPLMVLVIDYLIVPAKSKSTLN 359
QY 345 HPYVLTAPWYINFIFFIQPHKEERFLFPVYPLICLGAVALSALOKCY----HFVFOR 400
DB 360 FPHYISLAPLYLLVLLVFAQPHKEERFLFPVYPLISLCAITVDVYVYFFRMKSVYVFKI 419
QY 401 YRLRLEHYVTWGNALGTVFLFGLSFSRVALFRGVHGPLDLPEFYRI-AT---DPTTH 456
DB 420 KAGVYLDHSMFIALVMVSTLLGLSRVAFALYRNHAPMDLMLNLNQKATPQYDPDV- 478
QY 457 TVPEGRPVNVCWKWYRFPSSFLP-DNKQLOIFSEFRGOLPKPPAEGPLATRIPTDMND 515
```

```

b 479 -----IYNVCIGKDWHRYPGSPFPKAKNFRFLKXGEFRGMLPAYVDEGONATKVQVPY 532
y 516 MNDQNLBPSRIDISKHYLVLDITMRETPREPKYSSNKEWISLAYRFLDASRSLK 575
b 533 FNDLNOENHMYFDYDRCDLFDVDEGKYTALENYKRSKSDMSVMSKSLFLLPEKSHKV 592
y 576 LRAFYVPELSDOYTVVYNTYIKPRKAKO 604
b 593 LRAFYVPELTDNHIQYGFNLK-RKTKR 620

RESULT 9
9BGVO
D C9BGVO PRELIMINARY; PRT; 198 AA.
C C9BGVO;
T 01-JUN-2001 (TREMELrel. 17, Created)
T 01-JUN-2001 (TREMELrel. 17, Last sequence update)
T 01-DEC-2001 (TREMELrel. 19, Last annotation update)
E Hypochemical 22.0 kDa protein.
S Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
C Cercopithecinae; Macaca.
X NCBI_TaxID=9541;
N [1]
P SEQUENCE FROM N.A.
C TISSUE=frontal cortex;
A Oseada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
A Suzuki Y., Sugano S., Hashimoto K.;
T "Isolation of full-length cDNA clones from macaque brain cDNA
T libraries.";
L Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; AB055306; BAB21931.1; -.
W Hypochemical protein.
Q SEQUENCE 198 AA; 22001 MW; 9276CF0AD4F3F82 CRC64;

Query Match 30.5%; Score 997; DB 6; Length 198;
Best Local Similarity 95.9%; Pred. No. 5.1e-73;
Matches 188; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

y 1 MASRGARQRLKSGASSGTAAADKLRELLSGRAGAEHRTLSGNKAGQVWAPGSGT 60
b 1 MASRGARQRLKSGSSGTAAADKLRELLSGRAGAEHRTLSGNKAGQVWAPGSGT 60
y 61 AFKLLSARLCAALLSNISDCDETNYWEPHYLYIGEGFTWYSPAYAIRSYAYLLH 120
b 61 AFKLLSARLCAALLSNISDCDETNYWEPHYLYIGEGFTWYSPAYAIRSYAYLLH 120
y 121 AWPAAFAHRILOQNKILVYFLRCLLAFVSCICELYFYKAVCKKFGHVSMMFLAVLS 180
b 121 AWPAAFAHRILOQNKILVYFLRCLLAFVSCICELYFYKAVCKKFGHVSMMFLAVLS 180
y 181 TGMFCSSSAFLPSSFC 196
b 181 TGMFCSSSVVPSPLIC 196

RESULT 10
18LRHS
D C8LRHS PRELIMINARY; PRT; 596 AA.
C C8LRHS;
T 01-OCT-2002 (TREMELrel. 22, Created)
T 01-OCT-2002 (TREMELrel. 22, Last sequence update)
T 01-MAR-2003 (TREMELrel. 23, Last annotation update)
E OSJNBA0016109.26 protein.
N OSJNBA0016109.26.
S Oryza sativa (japonica cultivar-group).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
C Ehrhartoideae; Oryzaceae; Oryza.
X NCBI_TaxID=39947;
N [1]
P SEQUENCE FROM N.A.

```

```

RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone:OSJNBA0016109.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR Gramene; Q8LRHS; -.
DR SEQUENCE 596 AA; 67812 MW; 839AE12C2DF32292 CRC64;

Query Match 25.9%; Score 849.5; DB 10; Length 596;
Best Local Similarity 34.1%; Pred. No. 1.7e-60;
Matches 217; Conservative 84; Mismatches 249; Indels 87; Gaps 16;

QY 1 MASRGARQRLKSGAS-----SGTAPAAADKLRELLSGRAGAEHRTLSGNKAGQVWAP 56
DB 1 MSLSSARQR-RATAASPTDDGYSKEAKORRRRPSGEEBEGIR-----W---44
QY 57 EGSTAFKCLLSARLCAALLSNISDCDETNYWEPHYLYIGEGFTWYSPAYAIRSYAY 116
DB 45 --FLPFLALGLLRHWSASSNLHDCDEVFNWEPHYLYIGEGFTWYSSNPFALRSYLY 102
QY 117 LLLHANPAFAHRILOQNKILVYFLRCLLAFVSCICELYFYKAVCKKFGHVSMMFLAF 176
DB 103 LFIHALVAGPASMIFGEHKRVFVSFRIFLGLISTTETVLVVALSRKYGKRLACVVLAM 162
QY 177 LVLSTGMFCSSSAFLPSSFCWYTTLIANTGMVMDKTSIAVLGVAAGAILGWPFSAALGLP 236
DB 163 LCLTSGCFEAST-FLPSFSFMYATLSSALFLENAVAASVAAAGVILGWPFSLVFLP 221
QY 237 IAFDLVNMKHWKSFHWSMALILFLVPVWVVDISYYGKLVIAPIINVLVNTVPHGPD 296
DB 222 ITVYSLI-RGSFRVFLSGFLTSNMLLVLSVIADYCYGKWTASVFNLLKYNVFGGESH 280
QY 297 LYGTPEWYFYLINGFLNFVAFAALALVLPLTSLMEYLLORFHVQNLGHPYMLTAPMYI 356
DB 281 LYGTGSPSYFNGFNFNFAFILLFLGFVP----PARKYVPLD----LIVSPVYL 332
QY 357 WFIIFFIQHPKEERFLPVVPLICLCAVALSALQKCYHFVORYLEHVTVS--NWLAL 415
DB 333 WLAFLMSLQAKHEERFLVPIVPLICVAAAVIDPDPDFH---DKYSSEQSIAGLRLPLIL 389
QY 416 GTVFLFGLLSFSRVALFRGVHGLDLYPEFYRIATDPTIHTVPEGRFNVNVCYKWEYRF 475
DB 390 GPIL---CASHRTFSMLNGYCAPIYQHLE-----HHEDTGGSVLCVGSWEHY 438
QY 476 PSSFLP---DNMQLQFIPSEFRGQLKPFABGPLATRIPTMDNDQNLBPSRYI---528
DB 439 PSSFPPIPSVISEVM---WIDGFRGLLPFPFNETLGGTTAAPSFYFNDKKNKASDEQYVFL 495
QY 529 -----DISKCHYLVDLDMRETPREPKYSSNKEWISL 561
DB 496 APLKYSTVFLITLIRVVTSLHFWLQKDIACSLVVELDLRRFPSPR---GSLSTWEITL 552
QY 562 AYRPFELDASRSLKLRAPYVPLSDQYTVVYNTYILK 598
DB 553 AALPFLDRELSPALYRSFPFHRWQNNVFLGYKLLR 589

RESULT 11
Q9FZ49
ID Q9FZ49 PRELIMINARY; PRT; 570 AA.
AC Q9FZ49;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE F611.10 protein (Ser/Thr protein kinase, putative).
GN F611.10 OR AT1G16900.
OS Arabidopsis thaliana (Mouse-ear cress).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
C eurosids II; Brassicales; Brassicaceae; Arabidopsis.
X NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.

```



```
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Hallsworth K.;
RT "The sequence of C. elegans cosmid T27F7.";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Oriol R., Martinez-Duncker I., Chantret I., Mollicone R., Codogno P.;
RT "Common origin and evolution of glycosyltransferases using Dol-P-
monosaccharides as donor substrate.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58762; AAK39304.2; -.
DR EMBL; AJ431373; CAD24083.1; -.
DR WormPep; T27F7.3a; CE30458.
DR InterPro; IPR005599; PMP.
DR Pfam; PF03901; PMP; 1.
KW Hypothetical protein; Glycosyltransferase; Transferase.
SQ SEQUENCE 496 aa; 57191 MW; 6A0C53E5346D7F8C CRC64;

Query Match      8.1%; Score 264.5; DB 5; Length 496;
Best Local Similarity 24.4%; Pred. No. 3.9e-13;
Matches 102; Conservative 71; Mismatches 156; Indels 89; Gaps 19;

2Y 62 FKCLLSARLCAALLSNISDCDENWYWEPTHTLIYGEFGOTWEYSPAYAIRSYAYLLIHA 121
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 14 PRC-LSVYLVASWFPV---DEVYQSAEVAHVLVYGTGHLSEW--RHSLSRFFHPALIA 66

2Y 122 WPAAFHARILQTNKILVYFURCLLAFVSCICELYFYK---AVCKKFGHLVSRMMLAFIV 178
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 67 VLIKFDILSLSSOFLIYHIPRLAHALLFALADFSFYKICRLCKTKGI-AENSFVYLS 125

2Y 179 LSTGMFCSSAFPLSPFCMYTTLIAMTGWYMDKTSIAVLGVAAGAILG--WPFSALGI- 235
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 126 SWFVFCAPRT-LNSLETSLTIAL-NWPFETK-----NFKGTWPY-IALGVYL 173

2Y 236 -----PIAFDLLV--MKHRWKS-----PFHWSLMALILFLVPVVVIDSYGKLVIAIP 281
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 174 TIIIRPTVALIWLFGVYHLYHSPTVRLIFQLVPLTLPILIVTLIDSWAYGTPTIPL 233

2Y 282 LNIVLVNYFTPHGPDLYGTEPWYFLNGFLNFVAEALALLVPLTSLMEXYLQRFHVQ 341
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 234 WNFQGFNV-VOGGSALFGVHPWYIIVSGI-----PAVLTVQMIPIIVGLLG----P 280

2Y 342 NLGHFMYLTLAPMYIWFIFTIOPHKEERFLFPVYPLICLCGAVALSALCKYHFVFOR 401
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 281 NIFRPSLLPFPATITYIVHSLPHKEQRFLLPIPLLCIYAGGAFQNLK----- 331

2Y 402 RLEHYTVTSNW---LALGTVFLGLLSFS-----RSVALRFGYHGGLDL 442
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 332 -----WRGSAMVMVAINIGIALFTSRVHQVGFPTAPRIMEWRHGKLSV 379
```

Search completed: January 13, 2004, 07:03:17

Job time : 67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 13, 2004, 06:40:49 ; Search time 22 Seconds
(without alignments)
1175.088 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGASSGDT.....VNTILKPKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PTCUTS-COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	131.5	4.0	524	4	US-09-198-452A-52
2	114	3.5	687	3	US-08-834-467-2
3	114	3.5	687	3	US-09-396-177-2
4	113.5	3.5	597	4	US-09-328-352-4703
5	113	3.5	866	4	US-09-134-001C-4930
6	112	3.4	420	4	US-08-795-876-33
7	111	3.4	940	4	US-09-328-352-8165
8	110.5	3.4	423	4	US-08-795-876-38
9	110	3.4	316	4	US-09-549-848B-33
10	110	3.4	538	4	US-09-252-991A-23060
11	109.5	3.3	695	1	US-08-487-886-2
12	109.5	3.3	695	3	US-08-482-855-2
13	109.5	3.3	695	4	US-08-474-986-2
14	108	3.3	638	1	US-08-295-814E-13
15	108	3.3	638	1	US-08-240-783B-2
16	108	3.3	638	3	US-09-084-813-2
17	108	3.3	638	3	US-09-343-361-13
18	108	3.3	638	5	PCT-US92-09662-2
19	107	3.3	390	4	US-09-679-279-5
20	105.5	3.2	436	4	US-08-795-876-2
21	105.5	3.2	692	3	US-07-757-342D-6
22	105.5	3.2	1956	3	US-08-843-417-10
23	105.5	3.2	1956	3	US-09-527-013-10
24	104.5	3.2	422	4	US-09-625-188-12
25	104.5	3.2	3169	2	US-08-477-451-6
26	103.5	3.2	418	4	US-09-252-991A-17598
27	103	3.1	370	3	US-08-513-974B-26

28	103	3.1	370	3	US-08-513-974B-323	Sequence 323, Appl
29	103	3.1	370	3	US-09-172-353-5	Sequence 5, Appli
30	103	3.1	370	3	US-08-776-971-21	Sequence 21, Appl
31	103	3.1	370	3	US-08-776-971-104	Sequence 104, Appl
32	103	3.1	370	4	US-09-799-955-5	Sequence 5, Appli
33	103	3.1	370	4	US-09-461-436B-26	Sequence 26, Appl
34	101.5	3.1	418	3	US-09-030-267-5	Sequence 5, Appli
35	101	3.1	3200	2	US-08-477-451-8	Sequence 8, Appli
36	100.5	3.1	739	4	US-09-134-001C-3586	Sequence 3586, Ap
37	100	3.1	373	4	US-09-622-439-6	Sequence 6, Appli
38	100	3.1	439	4	US-09-107-532A-4168	Sequence 4168, Ap
39	99	3.0	373	4	US-09-622-439-26	Sequence 26, Appl
40	99	3.0	764	3	US-07-741-453A-54	Sequence 54, Appl
41	99	3.0	764	3	US-07-741-453A-60	Sequence 60, Appl
42	99	3.0	764	3	US-07-741-453A-61	Sequence 61, Appl
43	98.5	3.0	533	4	US-09-252-991A-21407	Sequence 21407, A
44	98.5	3.0	3031	1	US-07-689-008-2	Sequence 2, Appli
45	98	3.0	446	2	US-08-672-814D-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-09-198-452A-52
; Sequence 52, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment thereof and uses thereof, in particular for the diagnosis, prev
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prev
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 52
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...524
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-52

Query Match	4.0%	Score 131.5;	DB 4;	Length 524;
Best Local Similarity	24.1%	Pred. No. 5.3e-05;		
Matches	74;	Conservative 41;	Mismatches 103;	Indels 89; Gaps 17;
QY	107	PAYAIRSVAYLLHAWPAAPHARILQTNKILVFYFLRCLLAFVSCICELVYFYKAVCKKFG	166	
DB	233	PSSEVGPPASLLIGTLPQNRLDLFRQGLSHLFAISGWHFSLCATTLMW---LCALLP	289	
QY	167	LHVRMMLAFVLVSTGMFCSSAFPLSSFCMYTTLAMT---GWYMDKTSIAVLGVAAAG	222	
DB	290	LKIKK-ILSFVILT-----SLSCIFPMLSVMSRWSISVITLLCFSNCFSGCSGLNRLGAG	343	
QY	223	AIL-----GWPFSAALGL---PIADLLVMKHWKSPF--HW-----	254	
DB	344	FILCSIFSPSPFTVLSFLATGILLFFPKISFLYTP--WTQFLSPFLPIRYILAMT	401	
QY	255	---SIMALILFLVPPVVIDSYGKLVIAPIANVLNVNTPHGDLYGTPEWYFVNLNGF	311	
DB	402	LALSUSAQLFTVLPIM---QYFGSL---PLEGLLYNLIVP-----FTILP---	440	
QY	312	LNFNVAVALALLVLPLTS-LMEVLLQRFHQVNLGHPV-----WLTLPAMYIW---	357	
DB	441	---IIVFLIATIIILPCSPITEALIQGF---LSHPWLNHPNLTLSFAPVPPWMLTLA	493	
QY	358	-FIIFFI	363	
DB	494	SLIIFFI	500	

RESULT 2

S-08-834-467-2

Sequence 2, Application US/08834467

Patent No. 608015

GENERAL INFORMATION:

APPLICANT: Albert, Vivian

TITLE OF INVENTION: Glycine Transporter

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.00

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/834,467

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 314572-109

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 687 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

S-08-834-467-2

Query Match 3.5%; Score 114; DB 3; Length 687;

Best Local Similarity 20.4%; Pred. No. 0.006;

Matches 97; Conservative 50; Mismatches 174; Indels 154; Gaps 20;

Y 84 TNYWEPHYLYG---EGFQWEYSPAYAIRSYAYLLHAWPAAPHARILQTNKILVY 140
b 302 TPYVVLTLFVRGVTLEG-----AFDGMWYLTFCQDKILEAKWGDAAQIFY 351
Y 141 FLRC-----LAFVSCICELY-----FYKAVCKKFGHLVS 170
b 352 SLACAWGGLITWASYNKFNHNCYRDSVIITNCATSVAGVFIPLSGFMANHLGVDVS 411
Y 171 RWM-----LAFVLSTGM-----FCSSSAFLPSSFCMYTLLIAM-----T 205
b 412 RVADHGPGGLAVAYPEALTLPLSPMLSLFFFMILLGLGTQFCLETLVTAIVDEGN 471
Y 206 GWYMDKTSIAVLGVA-AGAILGWPFSAALGLPIAFDILLVMKRWKSPFHWLSMALILFLV 264
b 472 EWILQKTYVTGLGVAAGVFLGLPLTSQAGI---YWLLMDNYAASF-----SLVVISC 522
Y 265 PVVVIDSYVYKGLVIAPLNIVLVNFTPHGPDLYGTEPWYFYLLNGFLNFVAFALALLV 324
b 523 IMCVALMYIGH-----RNVFQDIQMWLGPPELFPFQICWRP-----VSPAIIFFI 568
Y 325 LPLTSLMEYLLQRFVQNLGHPYMLTLAPMYITWFIIFZIQPKBERFLFPVYPPLICLGA 384
b 569 LVFTVIOYQPIYNNHQ---YFGW---AAVAGFLMAL-----SSVLCIPLYAMFRLCRT 616

QY 385 VALSALQKCHTFVQRVLEHYTTS-NWALQTVLFGLLSFSRSVALFRYHGPDLILY 443
Db 617 DGDTLQ-----RLKNATKPSRDW-----GPALE 641
QY 444 PEFYRIATDPTIHTVPEGRPNVVCVGEWYRFPSSFLLPDNWLOQFIPSEFRGQL 498
Db 642 HRTGRYA--PTIAPSPED-----GFEVQSLHFDKQAPIVGVNGSRL 682

RESULT 3

US-09-396-177-2

Sequence 2, Application US/09396177

Patent No. 6251617

GENERAL INFORMATION:

APPLICANT: Albert, Vivian

TITLE OF INVENTION: Glycine Transporter

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.00

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/396,177

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/834,467

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: 314572-109

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 687 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-396-177-2

Query Match 3.5%; Score 114; DB 3; Length 687;

Best Local Similarity 20.4%; Pred. No. 0.006;

Matches 97; Conservative 50; Mismatches 174; Indels 154; Gaps 20;

QY 84 TNYWEPHYLYG---EGFQWEYSPAYAIRSYAYLLHAWPAAPHARILQTNKILVY 140
Db 302 TPYVVLTLFVRGVTLEG-----AFDGMWYLTFCQDKILEAKWGDAAQIFY 351
QY 141 FLRC-----LAFVSCICELY-----FYKAVCKKFGHLVS 170
Db 352 SLACAWGGLITWASYNKFNHNCYRDSVIITNCATSVAGVFIPLSGFMANHLGVDVS 411
QY 171 RWM-----LAFVLSTGM-----FCSSSAFLPSSFCMYTLLIAM-----T 205
Db 412 RVADHGPGGLAVAYPEALTLPLSPMLSLFFFMILLGLGTQFCLETLVTAIVDEGN 471
QY 206 GWYMDKTSIAVLGVA-AGAILGWPFSAALGLPIAFDILLVMKRWKSPFHWLSMALILFLV 264
Db 472 EWILQKTYVTGLGVAAGVFLGLPLTSQAGI---YWLLMDNYAASF-----SLVVISC 522
QY 265 PVVVIDSYVYKGLVIAPLNIVLVNFTPHGPDLYGTEPWYFYLLNGFLNFVAFALALLV 324

523 IMCVAMVYIGH-----RNYEQDQMGFLFPFPLFFQICWEP-----VSPAILIFFI 568
325 LPLTSLMEYLLQRPHVONLQHPWLTILAPMYIWIIFFTQPHKBERFLFPVPLICLGA 384
569 LVFTVIOQITVNHQY---YPGW-----AVAIGFLMAL-----SSVLCPILYAMFLQRT 616
385 VALSALQKCYHVFQVRLEHYHTVTS-NMLAGTVFLFGLLSFVSVALFRGVHGPFLDIY 443
617 DGDFTLLO-----RLKNAKPSRDW-----GPALLE 641
444 PEFVRIATDTHTVPGRPVNVCKWEKVRFPSSFLPDNNQLQIPSEPRGOL 498
642 HRTGRYA--FTIAPSPED-----GPEVQSLHDKAQPIVIGSNGSRL 682

RESULT 4
US-09-328-352-4703
; Sequence 4703, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4703
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4703

Query Match 3.5%; Score 113.5; DB 4; Length 597;
Best Local Similarity 23.0%; Pred. No. 0.0054;
Matches 90; Conservative 71; Mismatches 142; Indels 89; Gaps 22;

95 IYGGFQWTSYPAIYRSYAYILLHAW-----PAAPHARILQTNKILVYFLR----- 143
92 VIGQVYQTG--ADSYLLFLTWLLFLWLRPNIGIFALICITSQTLFLPFRQTFWAEK 149
144 ---CLLAFVSCICELYFYKAVCKKFGHLHVSMMMLA-FLVLS-TGM--FCSSS--AFLPSS 194
150 PPYLFLPALNLLSVQW--ICQKYTALAFIFIANFAVISITGMIFLSSENLPYLISA 207
195 FCMTTTLIANTGWTNDKTS-----IAVLGVAAGAILGWPPSAALGLPIADLLVMKGR 247
208 F--FLGIIAFY-YFPNKKDDQ-CASLMAVLGVATITWL-----VDGINQLFK 251
248 WKSPPHWSMALILF---LVPVVVIDSVYVGKLVIAPLNVLVNVFTPHGPDLYGTSPW 303
252 DSNEFILLAGIIFTWFLAISYFLIKIFQSRFYIPLAI-----GAWLAGLAG 302
304 YFYLI---NGFLNPNVAFALALLVPLTSLMEYLLQRP-----HVNQLGHP--- 346
303 AFLTVFETISLIIGIIF-VAVATLLTKSQSYFIRQAFYCLFVSQTAFLFHLGSETDQ 361
347 -YMTLAPMYIWFIIFTQPHKBERFLFPVPLICGVALSALQKCYHVFQVRLEH 405
362 ILWVLIAGIFILCHSYFLKEH---WFFILQMLATYG-IAVYLLQMDHSLNSLTQT 416
406 Y--TVTNSNWLALGTVFLFG---LLSPFSRVAL 432
417 YLNLVLLNVLVFPSSVLLIGSKAVVSVKRSIFL 448

RESULT 5
US-09-134-001C-4930
; Sequence 4930, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4930
; LENGTH: 866
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4930

Query Match 3.5%; Score 113; DB 4; Length 866;

Best Local Similarity 16.1%; Pred. No. 0.011;
Matches 103; Conservative 115; Mismatches 201; Indels 220; Gaps 30;

131 LQTNKILVYFLRCLLAFVSCICELYFYKAVCK-KFGLHVS-----MMLAFLVLSTGMF 184
101 LNTNQILLMMNALIVSIKSTIALF---CTYLFYSYISKNIISLISFIFVMSPLY 155
185 CSSSAFLP-----SSFCMYTTLIANTGMYMDKTSIAVLGV 219
156 FRFTIYMPFSDTFIWLPLLGLIERLLRDFKAGLFIITVSLILINFFY-AYFLLIIGI 214
220 A-----AGAILGWPFSAALGLPIADLLVMKHWKSFH----- 253
215 GYTLIRIYRHPKDSLTRWQASLTI---ICSALLALGNSMFVFFHGVQSFLNNRQSTG 271
254 ---W-----SLMALILF---VPVVVIDSVYVGKLVIAPLNVLVNVFTPH 293
272 QVNWIEHLNKDTNIFPDNYLIWIFLSIQALLTIKLYKHYFKLFAALLLTIIFAFILPF 331
294 GPDLYG---TEPYFYFLINGFLNFVAFALALIV-----LPLTSLMEYLLQRP--FHV 340
332 VDQLFNGFSAPQKRWHP-----ILAFNSILIGLFVKYFKTIRPKTYIYTNLIAGSVIYI 386
341 QNUGHY---WLTAPM--YIWFIIFTQPHKBERFLFPVPLICGVALSALQKCYH 395
387 SSISYNTFLPWLSPVWSVIGLILILIKERKVRVYLYLV-----SISIALSLMIT 439
396 FVFORVRL--EHYVTNSNWLALGTVFLGILLSFVSVAL----- 432
440 FVFIKQIIFQDHIINRANKRYINSSFYNSAIQSRSLVKEMQTKNDQDRIINRVRDEQNTIP 499
433 ---FRG-----YHGPDLYPEFYRIADPTHTVPEGRPVNVCKWEKRYPPSSFL 480
500 MYQNFKGLSTYSSIFHNLDFYDALKI--NLABESVSRVYQSTNARQNI-E-SLFSVKYL 556
481 LPDNWQLQITPSEF-----RGOLKPPFAEGPLATRIVPTDMNDONLEPSRYI 528
557 MKDYO-NTIPSYFKKVRGQVYIYKQLPLP-----SVKVTQNTYNHKSLLKP----- 605
529 DISKCHYLDV-----LDTNR-BTPREPKYSSNKEEWISLAYRPFLLDA 569
606 -IDREHAMIINGAIVTSKGTAYHSKVNLLDQTRVSTQNTITRYSNNE-----LTV 653
570 SRSSKLLRAFYPFLSDQYTVV-----NYTI 596
654 NKESGIKHLHPKNTIRDKYDFILTNIKRGDPDSNTYV 692

RESULT 6
US-08-795-876-33
; Sequence 33, Application US/08795876
; Patent No. 6403305
; GENERAL INFORMATION:
; APPLICANT: Gershengorn, Marvin C.
; APPLICANT: Geras-Raaka, Elizabeth
; APPLICANT: Nussenzweig, Daniel R.

TITLE OF INVENTION: STRATEGY TO CLONE DRUGS FOR G PROTEIN
TITLE OF INVENTION: COUPLED RECEPTORS
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,876
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: BRAMAN, SUSAN J.
REGISTRATION NUMBER: 34,103
REFERENCE/DOCKET NUMBER: 19603/1280
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-263-1636
TELEFAX: 716-263-1600
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
S-08-795-876-33

Query Match 3.4%; Score 112; DB 4; Length 420;
Best Local Similarity 21.5%; Pred. No. 0.0044;
Matches 90; Conservative 56; Mismatches 143; Indels 130; Gaps 25;
70 LCAALLSNISDCDETFN---YWEPP-----THYLI---YGEQFTWEYSPAY-----A 110
24 LCGVVSDYKDDDDKDLATLFEWDEKESGLTEYRLVSNKSPLOKQLPFIETSDASG 83
111 IRSYAVLLHAWPAAFHA-----RILOTN--KILVFYRLCLLAFVSCICELYFYKA 160
84 YLGINLRVLINWFISILAITGNIVILVITTSQYKLTVPFLMCLNLFAD-LC-IGIYLL 141
161 VCKKFGHVSRLMFLV-LSTGMFCSSAF---LPSSFCMYT-TLIAMTGWY----- 208
142 LIASVDIHTKSQVHNAIDWQTGAGCDAAGFTVPFASLSVTLTTLERWHTITHAMQ 201
209 MDKTSIAVLGVAAGAILGWPFSA-----LGLPIAFD-----LLVMKHR 247
202 LD-CRVLHRAASVMVGWIFAFALPFIPIGSIYMKVSCLEPMDISPLSQLYVN--- 257
248 WKSFFHWSLMALILFVPPVVVDSYYGKLVIAPIVLYLN-----VFTPHGP 295
258 -----SLVLNV-LAFVVICGVTHIYLTVRNPNIVSSSDTRIARMAFLIFT--- 305
296 DLYGTEPWYFYLINGFLNFWAFALALLVPL--TSLMEYLLQRPV-VONLGHFWLTLA 352
306 DFLCMAPISFFAIS-----ASLKPLITVSKAKILLVLFHPINSCANFLYAI- 353
353 PMYINFIPIOPHKEERLFPPVPLICLCAVALSALQKCYHVFQRYLHYTVTSN 411
354 -----FTKNFRDRDFI-----LLSKCG-----CYEMQAIQYRTSTSTVHN 389

RESULT 7
S-09-328-352-8165
Sequence 8165, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 8165
LENGTH: 940
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-8165

Query Match 3.4%; Score 111; DB 4; Length 940;
Best Local Similarity 20.6%; Pred. No. 0.021;
Matches 81; Conservative 57; Mismatches 154; Indels 102; Gaps 21;
QY 87 YWETHYLIYGEQFTWEYSPAYAIRSYAYLLHAWPAAFHARILQTNKILVYELRCLL 146
DB 499 HMTAIGWVAQGTALIVMGVTYRLSRYIGVIL-----VLLSLALFYQWANE 547
QY 147 AFVSCICELYFYKAVCKKFGHVSRLMFLVLTG-----MFCSSAFLPSSFCMYTTLI 202
DB 548 EFTLSTSIY-----ALAQFISAFYLLQYNSKEQRYFSASFSGIFLC----- 590
QY 203 AMTGWYMDKTSIAVLGVAAGAILG-----WPFSA-----ALGLPIAFDLV---MKHR 247
DB 591 -----LGNVAGAVAGVEIMAWHHALSPLYLMFAIAIAFSAIVHYKLRVQ 636
QY 248 WKSFFHWSLMALILFLV-----PVVV-----IDSVYVYKLVIAPIVLYNVVFTPHGP 296
DB 637 WQS-LQLLISLISLILVLGEAFASQVFTLFKWDLSLQQTFLVS---TILSGLFIMAQ 693
QY 297 -----LYGTEPWYFYLINGFLNFWAFALALLVPLT-SLMEYLLQRPVQNLGH--- 345
DB 694 SSHLGYVYKVMAGLSNLALAIWVAIFPKMPIVALAFVPPVYSLWAY---KAHKTTLLHQI 750
QY 346 PW-LTLAPWYINFIPIOPHKEERLF-PVYPLICLCAVALSALQKCYHVFQRYEL 403
DB 751 PWCLSL-----IWLIVVSDRHSAYLYFVPLNLTDFPSILVFGALL-----FIYQHAP 802
QY 404 EHYTVTSNWLALGTVFLFGLLSFSRSVALFRGYH 437
DB 803 DQ-DKSLWTFKITTILVGLLVFSSVVV--RGLH 833

RESULT 8
US-08-795-876-38
Sequence 38, Application US/08795876
Patent No. 6403305
GENERAL INFORMATION:
APPLICANT: Gershengorn, Marvin C.
APPLICANT: Geras-Raaka, Elizabeth
APPLICANT: Nussenzveig, Daniel R.
TITLE OF INVENTION: STRATEGY TO CLONE DRUGS FOR G PROTEIN
TITLE OF INVENTION: COUPLED RECEPTORS
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: USA
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,876
FILING DATE:
CLASSIFICATION: 530

```
; ATTORNEY/AGENT INFORMATION:
; NAME: BRAMAN, SUSAN J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 19603/1280
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-263-1636
; TELEFAX: 716-263-1600
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-795-876-38

Query Match      3.4%; Score 110.5; DB 4; Length 423;
Best Local Similarity 21.3%; Pred. No. 0.0065;
Matches 90; Conservative 56; Mismatches 143; Indels 133; Gaps 25;

QY 70 LCAALLSNISCDSTFN-----YNEP-----THYLI--YGEGFQWEYSPAY-----109
DB 24 LCOGVSDYKDDDDKLDATLWPFWEDEEKNESGLTYRLVNSINKSPQKQLPAFISED 83
QY 110 --AIRSYAYLLHAWPAAPHA-----RILOTN--KILVVFELRCLLAFVSCICELYF 157
DB 84 ASGYLGYNILRVLIWIFISILAITGNIIVLITTSQVKLTVPFLMCLNLAFAAD-LC-IGI 141
QY 158 YKAVCKEGLHVSRRMLAFIV-LSTGMFCSSAP---LPSFCMYT-TLIAMTGWY-----208
DB 142 YLLIASVDIHTKGOYHNYAIDWQTGAGCDAAGFTVFASELSYTYLTALTIERWHTITH 201
QY 209 ---MDKTSIAVLGVAAGAILGWPFSAA-----LGLPIAPD-----LLVM 244
DB 202 AMQLD-CKVQLRHAASVMVWGCIWIFAPAAALFPPIGISSYMKVSIICLPMDIDSPLSQLYM 260
QY 245 KIRWKSFFHNSLMALILFLVPPVVIDSYGYKLVAPLNVLYN-----VFTP 292
DB 261 -----SLLVNV-LAFWVCGCYHIYLVNPNIVSSSDTRIAKRMAMLIFT- 308
QY 293 HGPDLVGTPEPVYFYLYNGFLNFVAFALALVLPL--TSLMEYLQRFH-VONLGHYPYL 349
DB 309 ---DFLWAPISFFAIS-----ASLKVLITVSKAKILLVLFHPINSCANFFLY 354
QY 350 TLAEMYWPIFFIOPHKEERFLPPVPLICCAVALSALQCKYHFVORYLEHYTWT 409
DB 355 AI-----FTKNFRDRDFI-----LISKCG-----CYEMQAQIYRTSTSTV 390
QY 410 SN 411
DB 391 HN 392

RESULT 9
US-09-549-848B-33
; Sequence 33, Application US/09549848B
; Patent No. 6541259
; GENERAL INFORMATION:
; APPLICANT: Lasener, Michael
; APPLICANT: Post-Breitenmiller, Dusty
; APPLICANT: Savidge, Beth
; APPLICANT: Weiss, James
; TITLE OF INVENTION: Nucleic Acid Sequences Involved in
; TITLE OF INVENTION: Tocopherol Synthesis
; CURRENT APPLICATION NUMBER: US/09/549,848B
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/129,899
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/146,461
; PRIOR FILING DATE: 1999-07-30
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 33
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Synecchocystis sp
US-09-549-848B-33

Query Match      3.4%; Score 110; DB 4; Length 316;
Best Local Similarity 22.6%; Pred. No. 0.0046;
Matches 61; Conservative 43; Mismatches 86; Indels 86; Gaps 13;

QY 173 MLAPLVLTGMFCSSSAFLPSSFCMTTLIAMTGMVMDKTS-LAVLGVAAGAI---LGWP 228
DB 110 VLSPALMATTFNVVLSGCLASGVFY--MLVYTHWLKRHTAQNIIVGGAGSIPPLVGM- 166
QY 229 FSAALGLPIADPDLVLMKRWKSFHWSLMALILFLVP-----VVVIDSYGYKLVIA 281
DB 167 -AAVTG-----DL-----SWTPWVLEFALIPLWTPPHFWALALMIKDDY-----AQ 205
QY 282 LNIVLVNVFTPHGPDLYGTEPWYFYLINGFLNFVAFALALVLPLTSLMEYLQRFHVQ 341
DB 206 VNVMLPVIA--GEETKVSQIWIY-----SLLVFPFSLLVLPQLGI- 247
QY 342 NLGHPYLTLPAPYIMFIIPFIQPHKEER-----FLFPVPLICCAVALSAL 390
DB 248 -----LYLAIAIILGGQFLVKAQWQKQAPGDRDLARGLFKFSFYLMCLLAWVIDSL 300
QY 391 QKCVHFVFORVLEHYTTSNWLALGTVEL 420
DB 301 PVTHQLVAQ-----MGTLILL 315

RESULT 10
US-09-252-991A-23060
; Sequence 23060, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23060
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23060

Query Match      3.4%; Score 110; DB 4; Length 538;
Best Local Similarity 22.9%; Pred. No. 0.011;
Matches 75; Conservative 39; Mismatches 96; Indels 118; Gaps 20;

QY 167 LHVSRRMLAF---LVLSTGMFC---SSSAFLPSSFCMTTLIAMTGMVMDKTSIAVLGVA 220
DB 123 LHTNAVIFAFGGCAIFATSYVYVQRTSQARLISD-----TLAFTFWGQ-----AVI 170
QY 221 AGAILGWP--FS-----AALGLPIADPDLVLMKRWKSFHWSLMALILF-----LVPV 267
DB 171 VGAVLTLPQGFTTSKEYAELEWPLALLAIV-----WITYAIVFGTIVKRVKXHI 221
QY 268 VIDSYYGK--LVIAPLNV-----LYNVTPHGPDLYGTEPWY-----FYLI 308
DB 222 YVGNWFYGAFLVLTAMLHVNHMSLFVSWFKSYSAYS--GATDAMVQWVWYGHNAVGFELT 279
QY 309 NGFLNENVAFALALLVPLTSLMEYLQRFHVQNLGHYPVWLTLPAPYIMFIFFIOPHKE 368
DB 280 TGFLGMWYFVFPKQARPVYS---YRLSIHV-----FW-ALISLIYW-----AGPH-- 321
```

Y 369 ERPLFPVYPLICLGAVALSALQKCYHFVORYLEHYVTNSWL-ALGTVFLFGLISFS 427
 b 322 -----HL-----HYTALPDWAQSLGVMVMSLILAPS 347
 Y 428 RVALFRGVHGLDLYPEFYRIATPTI 455
 b 348 WQGM-----NGWMTLSGAWHKLRTDPIL 371

ESULT 11

S-08-487-886-2

Sequence 2, Application US/08487886

Patent No. 5744448

GENERAL INFORMATION:

APPLICANT: Kelton, Christie Ann
 APPLICANT: Schweickhardt, Rene Lynn
 APPLICANT: Cheng, Shirley Vui Yen
 APPLICANT: Nugent, No. 5744448een Patrice
 TITLE OF INVENTION: Human Pollicle Stimulating
 TITLE OF INVENTION: Hormone Receptor
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Stephan P. Williams,
 ADDRESSEE: Ares-Serono, Inc.
 STREET: Exchange place, 37th floor
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density

COMPUTER: IBM PS/2, model 55 SX

OPERATING SYSTEM: MS-DOS version 4.0

SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,886

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/670,085

FILING DATE: 15-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Stephan P.

REGISTRATION NUMBER: 28546

REFERENCE/DOCKET NUMBER: US/252

TELEPHONE: (617) 723-1300

TELEFAX: (617) 723-8923

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 695

TYPE: Amino acid

TOPOLOGY: Linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: signal sequence

LOCATION: -17 to -1

IDENTIFICATION METHOD: hydrophobic

FEATURE:

NAME/KEY: putative amino-terminal extracellular domain

LOCATION: 1 to 349

IDENTIFICATION METHOD: similarity with other

IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular

FEATURE:

NAME/KEY: transmembrane domain

LOCATION: 350 to 613

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains

FEATURE:

NAME/KEY: putative transmembrane region I

LOCATION: 350 to 370

IDENTIFICATION METHOD: similarity to other G

FEATURE:

NAME/KEY: putative transmembrane region I

LOCATION: 350 to 370

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

NAME/KEY: putative transmembrane region II

LOCATION: 382 to 404

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region III

LOCATION: 427 to 448

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region IV

LOCATION: 469 to 491

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region V

LOCATION: 512 to 533

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region VI

LOCATION: 557 to 580

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region VII

LOCATION: 592 to 613

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative carboxy-terminal intracellular

NAME/KEY: domain

LOCATION: 614 to 678

US-08-487-886-2

Query Match 3.3%; Score 109.5; DB 1; Length 695;

Best Local Similarity 22.2%; Pred. No. 0.018;

Matches 81; Conservative 48; Mismatches 125; Indels 111; Gaps 21;

QY 105 YSPAVAIRSYAVLLHAWPAAFHA-----RIQTN--KILVYFELCLLAFVSCICE 154

Db 353 FNPCEIDMGYNILRVLIWFISILAITGNIIIVLLITTSQYKLTVPFELMCNLAFAAD-LC- 410

QY 155 LYFYKAVCKKFGHVSRRMLAFV-LSTGMFCSSAF---LPSSFCWYT-TLIAMTQWY- 208

Db 411 IGIYLLIASVDIHTKSOYHNYAIDWQTGACDAAGPFTVFASLSVYTLTTLERWHT 470

QY 209 -----MDKTSIAVLGVAAGAILGWPSAA-----LGLPIAFD-----L 241

Db 471 ITHAQOLD-CKVQLNHAASVVMGMWIFAFAPALFPFGISSYKVSICLPMDIDSPLSQL 529

QY 242 LYWKHRWKSFFHWSLMALILEFVPVWVIDSYYGKLVIAPIINVLYN-----V 289

Db 530 YW-----SLLVLN-LAFWVICGCIYHILTVRNPNIVSSSSDTRIARKRAMLI 578

QY 290 FTFPHGPDLYGTEPWYFYLLNGFLNPNVAFALALVLP--TSLMEYLLQRFH-VQNLGHP 346

Db 579 FT-----DFLCWAPISFFALS-----ASKVPLITVSKAKILLVLFHPINSCNP 623

QY 347 YWLTLPAPYINFIPIQHPKEERFLFPVVPPLICGVALSALQKCYHFVORYLEHY 406

Db 624 FLYAI-----FTKNPERDFFI-----LLSKCG-----CYEMQAQIYRTETS 659

QY 407 TWTSN 411

Db 660 STVHN 664

RESULT 12

US-08-482-855-2

Sequence 2, Application US/08482855

Patent No. 6121016

GENERAL INFORMATION:

APPLICANT: Kelton, Christie Ann

APPLICANT: Schweickhardt, Rene Lynn

APPLICANT: Cheng, Shirley Vui Yen

APPLICANT: Nugent, No. 6121016een Patrice

TITLE OF INVENTION: Human Follicle Stimulating

TITLE OF INVENTION: Hormone Receptor

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Stephan P. Williams,

ADDRESSEE: Ares-Serono, Inc.

STREET: Exchange Place, 37th floor

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density

COMPUTER: IBM PS/2, model 55 SX

OPERATING SYSTEM: MS-DOS version 4.0

SOFTWARE: VAX/VMS Mas11 via Kermit to IBM MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/482,855

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/670,085

FILING DATE: 15-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Stephan P.

REGISTRATION NUMBER: 28546

REFERENCE/DOCKET NUMBER: US/252

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 723-1300

TELEFAX: (617) 723-8923

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 695

TYPE: Amino acid

TOPOLOGY: Linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: signal sequence

LOCATION: -17 to -1

IDENTIFICATION METHOD: hydrophobic

FEATURE:

NAME/KEY: putative amino-terminal extracellular domain

LOCATION: 1 to 349

IDENTIFICATION METHOD: similarity with other

IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular

IDENTIFICATION METHOD: domains, hydrophilic

FEATURE:

NAME/KEY: transmembrane domain

LOCATION: 350 to 613

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains

FEATURE:

NAME/KEY: putative transmembrane region I

LOCATION: 350 to 370

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region II

LOCATION: 382 to 404

FEATURE:

IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:

NAME/KEY: putative transmembrane region III

LOCATION: 427 to 448

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region IV

LOCATION: 469 to 491

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region V

LOCATION: 512 to 533

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region VI

LOCATION: 557 to 580

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative transmembrane region VII

LOCATION: 592 to 613

IDENTIFICATION METHOD: similarity to other G

IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,

IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length

FEATURE:

NAME/KEY: putative carboxy-terminal intracellular

NAME/KEY: domain

LOCATION: 614 to 678

US-08-482-855-2

Query Match 3.3%; Score 109.5; DB 3; Length 695;

Best Local Similarity 22.2%; Pred. No. 0.018;

Matches 81; Conservative 48; Mismatches 125; Indels 111; Gaps 21;

QY 105 YSPAYAIRSYAYLLHAWPAFAH-----RILOTN--KILVFYFLRCLLAFVSCICE 154

Db 353 PNPCEIDMGYNILRVLIWFI-SILAITGNIIVILVITTSQYKLTVPFRLMCLNLAFAAD-LC- 410

QY 155 LYFYKAVCKKFGHLVSRMLAFIV-LSTGMFCSSAF---LPSSFCMYT-TLIAMTGWY- 208

Db 411 IGIYLLIIASVDIHTKSQYHNYAIDWQTGAGCDAAAGFTTVFASLSVYTLTATTLERWHT 470

QY 209 -----MDKTSIAVLGVAAGAILGNPFSA-----LGLPIAFD-----L 241

Db 471 ITHAQILD-CKVQLRHAASVVMVGMWIFAFPAALFPFEGISSYMKVSIKLPMDIDSPSQL 529

QY 242 LVMKRWKSPFHHSLMALIILFVFPVVVIDSYGKLVIAPIPLNVLVN-----V 289

Db 530 YVM-----SLLVLNV-LAFVVICGYIHILTVRNPINVSSTSDTRIAKRVMALI 578

QY 290 FTTPHGPDLGTEPWYFYVYLINGFLNFVNAPALALLPL--TSLMEYLLQRFH-VOMLGH 346

Db 579 FT----DFLCNAPISFFAIS-----ASLKVPLITVSKAKILLVLVLFHPINSCAP 623

QY 347 YWLTAPMYIWFIFFIQPHKEERFLFPVYPLILCLGAVALSALQKCYHFVFORYLEHY 406

Db 624 FLYAI-----FTKNFRROFFI-----LLSKCG-----CYEMQAQIVRTETS 659

QY 407 TVTSN 411

Db 660 STVHN 664

RESULT 13

3-08-474-986-2

LOCAL INFORMATION:
 APPLICANT: Kelton, Christie Ann
 Schweickhardt, Rene Lynn
 Cheng, Shirley Vui Yen
 Nugent, No. 63721leen Patrice
 TITLE OF INVENTION: Human Follicle Stimulating
 Hormone Receptor

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan P. Williams,
Ares-Serono, Inc.
STREET: Exchange Place, 37th floor

STREET: Exchange Place, 37th Floor
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109

COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
 COMPUTER: IBM PS/2, model 55 SX
 OPERATING SYSTEM: MS-DOS version 4.0
 SOFTWARE: VAX/VMS Maselli via Kermit to IBM MS-DOS

SEQUEN
5-08-474-986-2

Query Match	3.3%;	Score 109.5;	DB 4;	Length 695;
Best Local Similarity	22.2%;	Pred. NO. 0.018;		
Matches	81;	Conservative	48;	Mismatches 125;
				Indels 111;
				Gaps 21;

[illegible]

RESULT 14

US-08-295-814E-13
Sequence 13, Application US/08295814E
Patent No. 5658786
GENERAL INFORMATION:
APPLICANT: Smith, Kelli E.
APPLICANT: Borden, Laurence A.
APPLICANT: Hartig, Paul R.
APPLICANT: Weishank, Richard L.
TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/295,814E
FILING DATE: DECEMBER 19, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40558-B-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-295-814E-13

Query Match	3.3%; Score 108; DB 1; Length 638;
Best Local Similarity	20.4%; Pred. No. 0.023;
Matches	75; Conservative 45; Mismatches 144; Indels 104; Gaps 15;
QY	84 TENWEPHYLIYG---EGQQTWEYSPVAIRSAYVLLHAWPAAPHARILQTNKILVY 140
DB	253 TTPYVVLTLFVRGVTLG-----AFTGIMNYLTPKWDKILEAKWGDAAASQIFY 302
QY	141 FLRC-----LAFVSCICELY-----FYKAVCKFGELHVS 170
DB	303 SLGCAGWGGLITMASYNKFNHNCYRDSVIISITNCATSVYAGVFISILGFMANHLGLVDVS 362
QY	171 RMV-----LAFVLVSTGM-----FCSSSAFLPSRCQMYTLIAM-----T 205
DB	363 RVADHGPGAFVAYPEALTLLPISLWLSFFPMLILGLGTQFCLELFTVAIVDEWGN 422
QY	206 GWYMDKTSIAVLGVA-AGAILGWPFESAALGLPIAFDLLVMKHWKSFPHWSLMALILFLV 264
DB	423 EMILQKTYVTILGVAVAGFLGLTFSQAGI---YWLMLMDNYAASF-----SLVVISC 473
QY	265 PVVWIDSYYYGKLVIAPLNIVLVNFTPHGPDLYGTEPWYFYLIINGLFNFWAFALALLV 324
DB	474 IMCVSITMYIGH-----RNYFQDIQMWLGPPPLPFFQICRWFSPTIIF---FI 519
QY	325 LPLTSLMEVLLQRHVQNLGHPPWYLTAPWYIWFILFFIQPHKEERFLPVVPYPLICOGA 384
DB	520 LFTVIOVRPITYNHQ---YPGV-----AVAGFLMAL-----SSVICPLTALFQLCRT 567
QY	385 VALSALQK 392

Db 568 DGDILLOR 575

RESULT 15

US-08-240-783B-2
; Sequence 2, Application US/08240783B
; Patent No. 5/56348
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Branche, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinschenk, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240.783B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-240-783B-2

Query Match 3.3%; Score 108; DB 1; Length 638;
Best Local Similarity 20.4%; Pred. No. 0.023;
Matches 75; Conservative 45; Mismatches 144; Indels 104; Gaps 15;
QY 84 TNYWEPHYLIYG---EGFQWYSPAYAIRSYAYLLHAWPAAPFARILQTNKILVY 140
Db 253 TFPYVVLITLVRGVTEG-----AFTGIMYLTTPKWDKILEAKVWGDAASQIFY 302
QY 141 FLRC-----LLAFVSGICELY-----FYKAVCKKFGHLVS 170
Db 303 SLGCWAGGLITWASYNKHNNCYRDSVIISITNCATSVIAGVTFISILGFMANHLGVDVS 362
QY 171 RMW-----LAFVLSTGM-----FCSSAFSPSFCMYTTLIAM-----T 205
Db 363 RVADHGPGLAFAVPEALTLLPISLWSLLFFLFFMLILGLGTQCLLETIVTAIVDSVGN 422
QY 206 GYMDEKTSIAVQVA-AGAILGWPSAALGLPIAFDILLVMKRWKSPFHSMLAILFLV 264
Db 423 EWLQKTKYTVTLGVAVAGFLGIPITSOAGI---YWLMDNDNYAASF-----SLVVISC 473
QY 265 PVVVIDSVYYGKLVIAPLNIVLVNFTPHGPDLYGTEPWYFYLLINGFLNFNVAFAALLV 324
Db 474 INCVSMYIYGH-----RNVFQDIQMWLGFPPELFFQICWRVFSPTIIF---FI 519
QY 325 LPLTSLMEYLLQRFHVQNLGHFYWLTAPMYIWIIFFIQPHKEERFLFPVYPLICLGA 384

Db 520 LIFTVIQYRPITYNNHYQ---YPGW-----AVAIGFLMAL-----SSVICIPLYALFQLCRT 567
QY 385 VALSALOK 392
Db 568 DGDILLOR 575

Search completed: January 13, 2004, 07:03:43
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 13, 2004, 07:03:24 ; Search time 53 Seconds
(without alignments)
2323.065 Million cell updates/sec

Title: US-09-922-225A-2

Perfect score: 3274

Sequence: 1 MASRGARQLKSGASSGDT.....VNYTLKPKAKQIRKSGG 611

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747907 seqs, 201509753 residues

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3274	100.0	611	11	US-09-922-225A-2
2	1797	54.9	339	11	US-09-922-225A-8
3	1489	45.5	331	11	US-09-922-225A-6
4	1395	42.6	258	15	US-10-050-704-121
5	1358	41.5	623	11	US-09-922-225A-14
6	1013.5	31.0	255	11	US-09-922-225A-4
7	986	30.7	653	11	US-09-922-225A-13
8	973	29.7	577	11	US-09-922-225A-11
9	973	29.7	577	11	US-09-922-225A-15
10	812	24.8	570	11	US-09-922-225A-12
11	209.5	6.4	554	12	US-10-231-265-240
12	130	4.0	433	16	US-10-080-170-547
13	126	3.8	584	12	US-10-032-585-7898
14	117	3.6	440	16	US-10-080-170-164
15	111.5	3.4	458	9	US-09-862-767A-9

16	111.5	3.4	899	12	US-10-318-306A-48	Sequence 48, Appl
17	111.5	3.4	899	12	US-10-319-236A-48	Sequence 48, Appl
18	111.5	3.4	1210	9	US-09-922-217-692	Sequence 692, App
19	111.5	3.4	1210	10	US-09-833-263-692	Sequence 692, App
20	111.5	3.4	1210	14	US-10-035-380-692	Sequence 692, App
21	111.5	3.4	1548	9	US-09-922-217-1085	Sequence 1095, Ap
22	111.5	3.4	1548	14	US-10-035-380-1095	Sequence 1095, Ap
23	111	3.4	460	9	US-09-741-669-464	Sequence 464, App
24	110.5	3.4	695	12	US-10-349-838A-22	Sequence 606, App
25	110	3.4	314	12	US-10-017-161-606	Sequence 38, Appl
26	110	3.4	314	12	US-09-908-006A-38	Sequence 38, Appl
27	110	3.4	314	12	US-10-292-798-532	Sequence 532, Appl
28	110	3.4	216	12	US-10-349-508-33	Sequence 33, Appl
29	110	3.4	755	15	US-10-156-761-12016	Sequence 12016, A
30	109.5	3.3	215	10	US-09-747-155-421	Sequence 421, App
31	109.5	3.3	604	12	US-10-292-798-1170	Sequence 1170, Ap
32	109.5	3.3	695	10	US-09-804-626-8	Sequence 8, Appli
33	109.5	3.3	695	12	US-10-349-838A-2	Sequence 26, Appli
34	109.5	3.3	695	12	US-10-349-838A-26	Sequence 26, Appli
35	109.5	3.3	695	12	US-10-349-838A-28	Sequence 28, Appl
36	109.5	3.3	695	12	US-10-349-838A-30	Sequence 30, Appl
37	109.5	3.3	695	12	US-10-349-838A-34	Sequence 34, Appl
38	109.5	3.3	695	15	US-10-225-567A-122	Sequence 122, App
39	108.5	3.3	406	12	US-10-207-655-65	Sequence 65, Appl
40	108.5	3.3	406	12	US-10-091-007-12	Sequence 12, Appl
41	108	3.3	317	9	US-09-815-242-13638	Sequence 13638, A
42	108	3.3	633	12	US-10-205-219-30	Sequence 30, Appl
43	108	3.3	638	12	US-09-845-908-13	Sequence 13, Appl
44	107.5	3.3	695	12	US-10-349-838A-24	Sequence 24, Appl
45	107.5	3.3	695	12	US-10-349-838A-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-09-922-225A-2
; Sequence 2, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; TITLE OF INVENTION: Associated with Bipolar Disorder
; FILE REFERENCE: P-EA 4572
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-225A-2

Query Match 100.0%; Score 3274; DB 11; Length 611;
Best Local Similarity 100.0%; Pred No. 3.9e-309;
Matches 611; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MASRGARQLKSGASSGDTAPAADKLRELLSGREAGGAHRTLSGNKAGQVWPEGST	60
DB	1	MASRGARQLKSGASSGDTAPAADKLRELLSGREAGGAHRTLSGNKAGQVWPEGST	60
QY	61	AFKLLSARLCALLSNISDCDTEFNWETHYLLIYGEQFTWEYSPAYAIRSYALLLH	120
DB	61	AFKLLSARLCALLSNISDCDTEFNWETHYLLIYGEQFTWEYSPAYAIRSYALLLH	120
QY	121	AWPAAFHARILQTNKILVFLRCLAFVSCICELFYKAVCKKFGHLVSRMMLAFVLVS	180
DB	121	AWPAAFHARILQTNKILVFLRCLAFVSCICELFYKAVCKKFGHLVSRMMLAFVLVS	180
QY	181	TGMFCSSAFILPSSFCMWTTLIATGMVMDXTIAVLGVAAGAILGWPFSALGLPIAFD	240
DB	181	TGMFCSSAFILPSSFCMWTTLIATGMVMDXTIAVLGVAAGAILGWPFSALGLPIAFD	240

241 LLVVKHRKWSFFHSLMALIILFLVPVVVVIDSYTGKVLVIAPLNIIVLNVTTPHGPDLVGT 300
241 LLVVKHRKWSFFHSLMALIILFLVPVVVVIDSYTGKVLVIAPLNIIVLNVTTPHGPDLVGT 300
301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAPMYWFI 360
301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAPMYWFI 360
361 FTIQPKHERFLFPVYPLICLCGAVALSALOKCHYFVQRYLHYVTNSWALGTVEL 420
361 FTIQPKHERFLFPVYPLICLCGAVALSALOKCHYFVQRYLHYVTNSWALGTVEL 420
421 FGLLFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWYRFPSSFL 480
421 FGLLFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWYRFPSSFL 480
481 LPDNWLOQIPSEFRGOLPKPFAEGPLATRIPTDMNDQNLLEEPSRYIDISKHYLVLDL 540
481 LPDNWLOQIPSEFRGOLPKPFAEGPLATRIPTDMNDQNLLEEPSRYIDISKHYLVLDL 540
541 TWRETPREPKYSSNKEWISLAYRFLDASRSSKLLRAFYVFFLSDQYTVVNYTILKPR 600
541 TWRETPREPKYSSNKEWISLAYRFLDASRSSKLLRAFYVFFLSDQYTVVNYTILKPR 600
601 KAKQIRKXSGG 611
601 KAKQIRKXSGG 611

RESULT 2
S-09-922-225A-8
Sequence 8, Application US/09922225A
Publication No. US20030104385A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
FILE REFERENCE: P-EA 4672
CURRENT APPLICATION NUMBER: US/09/922,225A
CURRENT FILING DATE: 2003-01-14
NUMBER OF SEQ ID NOS: 117
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 339
TYPE: PRT
ORGANISM: Homo sapiens
IS-09-922-225A-8

Query Match 54.9%; Score 1797; DB 11; Length 339;
Best Local Similarity 100.0%; Pred. No. 6.4e-166;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MASRGARQLKSGASGGTAPADKRLRELLSREAGGAHRTLSGNKAGOWAPEGST 60
1 MASRGARQLKSGASGGTAPADKRLRELLSREAGGAHRTLSGNKAGOWAPEGST 60
61 AFKCLLSARLCAALLSNISDCDFTNYWEPHYLIYEGGFQWYSPAYAIRSYAALLH 120
61 AFKCLLSARLCAALLSNISDCDFTNYWEPHYLIYEGGFQWYSPAYAIRSYAALLH 120
121 AWPAAFHARILQNKILVFFLRCCLAFVSCICELYFYKAVCKKGLHVSRLMFLVLS 180
121 AWPAAFHARILQNKILVFFLRCCLAFVSCICELYFYKAVCKKGLHVSRLMFLVLS 180
181 TGMFCSSAFLPSFCMYTTLIATGWMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD 240
181 TGMFCSSAFLPSFCMYTTLIATGWMDKTSIAVLGVAAGAILGWPFSAALGLPIAFD 240
241 LLVVKHRKWSFFHSLMALIILFLVPVVVVIDSYTGKVLVIAPLNIIVLNVTTPHGPDLVGT 300
241 LLVVKHRKWSFFHSLMALIILFLVPVVVVIDSYTGKVLVIAPLNIIVLNVTTPHGPDLVGT 300

301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAP 339
301 EPWYFYLLNGFLNFNVAFAALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAP 339
RESULT 3
US-09-922-225A-6
Sequence 6, Application US/09922225A
Publication No. US20030104385A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
FILE REFERENCE: P-EA 4672
CURRENT APPLICATION NUMBER: US/09/922,225A
CURRENT FILING DATE: 2003-01-14
NUMBER OF SEQ ID NOS: 117
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 331
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-225A-6

Query Match 45.5%; Score 1489.5; DB 11; Length 331;
Best Local Similarity 89.3%; Pred. No. 4.9e-136;
Matches 284; Conservative 5; Mismatches 14; Indels 15; Gaps 3;

295 PDLXGTEFWYFYLLNGFL-NPNVAFALALLVPLTSLMEYLQRFHVQNLGHPYWLTLAP 353
28 PVLVLSKFW--KVDGGFTQNFN-----SRTEFCLMCVQNLGHPYWLTLAP 73
354 MYIMFIPIFIOPHKEERFLFPVYPLICLCGAVALSALOKCHYFVQRYLHYVTNSWL 413
74 MYIMFIPIFIOPHKEERFLFPVYPLICLCGAVALSALOKCHYFVQRYLHYVTNSWL 133
414 ALGTVFLFGLLSFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWY 473
134 ALGTVFLFGLLSFSRSVALFRGYHGPDLDPYEFYRIATDPTIHTVPEGRPVNVCVKWY 193
474 RFPSSFLPDWQLOQIPSEFRGOLPKPFAEGPLATRIPTDMNDQNLLEEPSRYIDISK 533
194 RFPSSFLPDWQLOQIPSEFRGOLPKPFAEGPLATRIPTDMNDQNLLEEPSRYIDISK 253
534 HYLVDLDMRETPREPKYSSNKEWISLAYRFPFLDASRSSKLLRAFYVFFLSDQYTVVYN 593
254 HYLVDLDMRETPREPKYSSNKEWISLAYRFPFLDASRSSKLLRAFYVFFLSDQYTVVYN 313
594 YTIKPRKAKQIRKXSGG 611
314 YTIKPRKAKQIRKXSGG 331

RESULT 4
US-10-050-704-121
Sequence 121, Application US/10050704
Publication No. US20030050442A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: 62 Human Secreted Proteins
FILE REFERENCE: P2039P1
CURRENT APPLICATION NUMBER: US/10/050,704
CURRENT FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: 09/684,524
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: PCT/US00/08979
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: 60/128,693
PRIOR FILING DATE: 1999-04-09
PRIOR APPLICATION NUMBER: 60/130,991
NUMBER OF SEQ ID NOS: 344
SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 121
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-050-704-121

Query Match
  42.6%; Score 1395; DB 15; Length 258;
Best Local Similarity 100.0%; Pred. No. 5.2e-127;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 MYWFIIFFIQPHKEERFLPVVPLICLCAVALSALQKCYHFVQRYRLEHYVTTSNKL 413
Db 1 MYWFIIFFIQPHKEERFLPVVPLICLCAVALSALQKCYHFVQRYRLEHYVTTSNKL 60

QY 414 ALGTVFLGGLSFSRSVALFGYHGPDLDYPEFYRIATDTTIHTVPEGRVNVVCVKWY 473
Db 61 ALGTVFLGGLSFSRSVALFGYHGPDLDYPEFYRIATDTTIHTVPEGRVNVVCVKWY 120

QY 474 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDNDQNLPEPSRYIDISK 533
Db 121 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDNDQNLPEPSRYIDISK 180

QY 534 HYLVDLDTWRETREPCKYSSNKEWISLAYRPFDASSKLLRAFYVPFLSDQYTYVYN 593
Db 181 HYLVDLDTWRETREPCKYSSNKEWISLAYRPFDASSKLLRAFYVPFLSDQYTYVYN 240

QY 594 YTIKPKRAKQIRKSGG 611
Db 241 YTIKPKRAKQIRKSGG 258

RESULT 5
US-09-922-225A-14
; Sequence 14, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 623
; TYPE: PRT
; ORGANISM: D. melangaster
US-09-922-225A-14

Query Match
  41.5%; Score 1358; DB 11; Length 623;
Best Local Similarity 42.4%; Pred. No. 7.6e-123;
Matches 267; Conservative 115; Mismatches 213; Indels 34; Gaps 10;

QY 1 MASRGARQRLKSGASSGDTAPADKXRLGLG---SREAGAEHRTLSGNKAGQVWA-- 55
Db 1 MAPPAARARYIAKADNOILSKPPRPGLGNKNTKEATPAGKKDKDAKKNQPTSGG 60

QY 56 -----PEGSTAFKLLSARLCAALISNIDCDETFNWEPTHTLYLVEGEGFTWEY 105
Db 61 QEKGLPNPMPVSQAKFTTFVSARLCSAIWAIADCDTFNWEPLHYIINGHGLQWEY 120

QY 106 SPAYAIRSYAILLHAWPAAFHARILQTNKILVFYFLRCLLAFVSCICELYKAVCKKF 165
Db 121 SFQGLRSYTYLLQGVPGYFYQKLFNPSILIFWVACMLGFGCAVMERYWYKSIQCF 180

QY 166 GLHVSMMLAFLVLTSGMFCSSAFLPSSFCNWTTLIAMTGMVMDKTSIAVLGVAAGAIL 225
Db 181 GHIGRLMLIFQVFSVGMFVSSTALLPSSFMFPCALAAWQNVCFALFLTAIGALL 240

QY 226 GWPFSAALGLPIAFDILLVMKRWKSFFFHWSLMALILFLPVVVVDSYVYKGLVIAPLNIV 285
Db 241 GWPFALIGIPVLEMLLRQDRWKTQVQWTLTSGATVAIPMAIDTSYFGKLTAPLNIV 300
```

```
QY 286 LYNVFTPHGPDLYGTBPWFYFLINGELNFENAFALALLVLPLTSLMEYLLQRFHYQN-LG 344
Db 301 WYNVFTSHGNIFGTBPLSYIINGELNFENIWLIA-LQLPIMLVIDYLIVPAKSKSTLN 359

QY 345 HPYWLTLAPMYIWIIFFIQPHKEERFLPVVPLICLCAVALSALQKCY----HFVQQR 400
Db 360 PPHVLSLAPLYLWLVFFAQPHKEERFLPVVPLISLCGAIITVDVYVFFRPMKSVFKI 419

QY 401 YRLEHYVTTSNMLAGTVFLGGLSFSRSVALFGYHGPDLDYPEFYRI-AT---DPTIH 456
Db 420 KAGVHYLDHSMFAILVMWVTSTLLGSRVFLYRNTAHMDLMLLELNQKATQYDQPDV- 478

QY 457 TVPEGRVNVVCVKWYRPPSSFLP-DNQLOFIPSEFRGOLPKPFAEGPLATRIVPTD 515
Db 479 -----LYNVICIGKWHRYPCSPFFPAKXFLRLKSEFRGMLPAYVDEGQNAIKVQVY 532

QY 516 MNDONLEPSRYIDISKHYLDLDTWRETREPCKYSSNKEWISLAYRPFDASSKLL 575
Db 533 FNDLNQENHMYFDYDCDFLVDDEGKYTALEPNYSKRSKOWSMKSLPFLIPEKSHKV 592

QY 576 LRAFYVPFLSDQYTYVYNYTIKPKRAKQ 604
Db 593 LRAFYVPFLTDNHQYGDENLLK-RKTKR 620
```

```
RESULT 6
US-09-922-225A-4
; Sequence 4, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922,225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-225A-4

Query Match
  31.0%; Score 1013.5; DB 11; Length 255;
Best Local Similarity 85.0%; Pred. No. 5.3e-90;
Matches 193; Conservative 5; Mismatches 14; Indels 15; Gaps 3;
```

```
QY 295 PDLYGTBPWFYFLINGEL-NFNVAFALALLVLPLTSLMEYLLQRFHYQNIGHYPWLTAP 353
Db 28 PVLVLSKFW--KVDGQGTQNFNL-----SRTEFCLWCVPQNLGHPYWLTLAP 73

QY 354 MYWFIIFFIQPHKEERFLPVVPLICLCAVALSALQKCYHFVQRYRLEHYVTTSNML 413
Db 74 MYWFIIFFIQPHKEERFLPVVPLICLCAVALSALQKCYHFVQRYRLEHYVTTSNML 133

QY 414 ALGTVFLGGLSFSRSVALFGYHGPDLDYPEFYRIATDTTIHTVPEGRVNVVCVKWY 473
Db 134 ALGTVFLGGLSFSRSVALFGYHGPDLDYPEFYRIATDTTIHTVPEGRVNVVCVKWY 193

QY 474 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDNDQNLPEPSRYIDISK 520
Db 194 RPPSSFLPDNQLOFIPSEFRGOLPKPFAEGPLATRIVPTDNDQNLPEPSRYIDISK 240
```

```
RESULT 7
US-09-922-225A-13
; Sequence 13, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
```

TITLE OF INVENTION: Associated with Bipolar Disorder
FILE REFERENCE: P-EA 4672
CURRENT APPLICATION NUMBER: US/09/922,225A
CURRENT FILING DATE: 2003-01-14
NUMBER OF SEQ ID NOS: 117
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 653
TYPE: PRT
ORGANISM: C. elegans
3-09-922-225a-13

Query Match 30.1%; Score 986; DB 11; Length 653;
Best Local Similarity 36.5%; Pred. No. 1.2e-86;
Matches 217; Conservative 99; Mismatches 213; Indels 66; Gaps 13;
Y 48 NKAGQVAPGEGTAKKLLSARLCAALLSNISDCDETFNWFPHYLIIYEGEGTWEYSP 107
b 43 NNPDNDWPFSGVFPMLLSIRISGAIIINDCEVYNYWEPLHLFLYEGEGTWEYSP 102
Y 108 AVAIRSYAVLLHAMPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFG 167
b 103 VTAIRSYFYIYLYPASILFANLPDGTIVVFTLRLTIGLFCLLGEVYAFDAICKINI 162
Y 168 HYSRMWLAFLVLTGMPGSSAFPLSPSCMTTLIAMTGMXTSTAVLGVAAGILGW 227
b 163 ATGRFFILFISFGMFLASTAFVSPSCMAITFYILGAYLNENWTAGIFCVAFSTWVG 222
Y 228 PPSAALGLPIADLLVMKHWKSPFHHLSMALILFLVPVVVIDSYGKLVIAPIPLV 287
b 223 PPSAVLGLPIADVMDLLKGLRFRILTSLVGLGIGGVQVITDSHYFGKVLAPLNI 282
Y 288 NYVTPHGPDIYTEPWYFYLINGFNVAFAALLVPLTSLMEYLLQRFVQNLGHPY 347
b 283 NYVSGPGLYGEELPSFYIKLNNVIFAAPFGPLS--LAYFTKVMMSQDENVAL 340
Y 348 WLTLAPMYI-----WFTIFPIQHKBERFLFPVYPLICLGAVALSALQCKVHF 400
b 341 YORFAPITLLAVTTAAWLLIFGSAHQKBERFLFPYIPFIAPFAALADATNR--- 397
Y 401 YLEHYTVTSNWLALGTFLFGLLSFSRSVALFRGYHGLDLYPEFYRIATDPT---- 456
b 398 LGWD-----NILSILFICFAILSASRTYSIHNNYGSVHEIYRSLNAELTNR 451
Y 457 TPEGEPVNVGVCKEYRPSPSFLPD-----NWQLQFIPSEFRGOLPKPFAEGPL 507
b 452 D-----PIRVCVGKWHRPSPSFFIPQTVSDGKKVEMRFQSEFRGLLPFLKSD 506
Y 508 ATRIVPTMNDONLEBPSRYIDISKCHYLVDLTWRETREPCKYSSNKEE----- 557
b 507 VTRHIPTENNLNQBEISRYVLDSDCYVDVD-MPQSDREPDFRQKRYKRTKKWK 565
Y 558 -----WISLAYR--PFLDASRSSKLLRAFYVFFLSDQ---YTVVYNYTLK 598
b 566 RIENAILQFWLNSLPEEIKFLNSN-----FHSFPFFVFDNFGLFQAYSHYSFLK 615

RESULT 8
3-09-922-225a-13
Sequence 11, Application US/09922225A
Publication No. US20030104385A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
FILE REFERENCE: P-EA 4672
CURRENT APPLICATION NUMBER: US/09/922,225A
CURRENT FILING DATE: 2003-01-14
NUMBER OF SEQ ID NOS: 117
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 577
TYPE: PRT

ORGANISM: S. cerevisiae
US-09-922-225a-11

Query Match 29.7%; Score 973; DB 11; Length 577;
Best Local Similarity 37.7%; Pred. No. 1.8e-85;
Matches 219; Conservative 94; Mismatches 204; Indels 64; Gaps 18;
QY 53 VWAPGEGTAFKCLLSARLCAALLSNISDCDETFNWFPHYLIIYEGEGTWEYSPAIR 112
DB 15 VW-----TFSILAVLRLTSASPRVIDDCDEVYNYWEPLHLLYGLQVWESPEYAIR 68
QY 113 SYAVLLHAMPAAFHARILQTNKILVFLRCLLAFVSCICELFYKAVCKKFGFLHVS 172
DB 69 SWFVIALHAFV-QFLARGGLSLRHLVFIYFIRGVLACFSACFETNLILAVARNFNRA 127
QY 173 MLAFVLITGMPGSSAFPLSPSCMTTLIAMTGMWY---MDKTSIAVLG-VAAGAIL 228
DB 128 LTSVLVFNSGWASATSFLESPFAMNVTILASQAQSPSTKRTKVVSFITIGAV 187
QY 229 PPSAALGLP-IAFDLLVMKHWKSPF-HWSLMALILFLVP--VVVIDSYGKLVIAPI 284
DB 188 FSAALSIPFILLELDVLDKGRFHLFCRWFKAFVALLITGICITVDSLFIYHRIQ 247
QY 285 VLVNVFTEP--HGPDLYTEPWYFYLINGFNVAFAALLVPLTSLMEYLLQRFHV 342
DB 248 VYVNLADGEGPDYGTETPWYFANLSLQHNIVLWFAWACGPLVLAAP----- 298
QY 343 LGHPYWTL-----APMYINFIPIQHKBERFLFPVYPLICLGAVALSALQ 392
DB 299 ---TNWINDSFLDSSVISGPFYIWLFIPIQHKBERFYPIYPLVCLAAAIGL 355
QY 393 CYHVFORYLEHYTVTS---NWLALGTFLFGLLSFSRSVALFRGYHGLDLYPEFY 448
DB 356 ---LMQILSINETVRSKFPVRFVLCVYVAILGCLSIARILAI-QYNAPMIY 411
QY 449 IATDPTIHTVPEGRPVNVGVCKEYRPSPSFLPDQWQLQFIPSEFRGOLPKPFA 507
DB 412 LETDNVTT-----NVCVGKWHRYSPSTFPLDNRSLKFKVSEDFGILPGEF 464
QY 508 -----ATRIVPTMNDONLEBPSRYIDISKCHYLVDLTWRETREPCKYSSNKE 560
DB 465 WNRREGYQIPEHNEFNNEEPTRYLSLESCDFLIDLEFDHSHKATVNEPIYSKS-D 523
QY 561 LAYRPFPLDASRSSKLLRAFYVFFLSDQTYVYNYTLKPKR 601
DB 524 VWYVFFIDTKTPEWGRAFAVPFTEPKWGRY-EILVKKPVK 563

RESULT 9
US-09-922-225a-15
Sequence 15, Application US/09922225A
Publication No. US20030104385A1
GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
FILE REFERENCE: P-EA 4672
CURRENT APPLICATION NUMBER: US/09/922,225A
CURRENT FILING DATE: 2003-01-14
NUMBER OF SEQ ID NOS: 117
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 577
TYPE: PRT
ORGANISM: S. pombe
US-09-922-225a-15

Query Match 29.7%; Score 973; DB 11; Length 577;
Best Local Similarity 37.7%; Pred. No. 1.8e-85;
Matches 219; Conservative 94; Mismatches 204; Indels 64; Gaps 18;
QY 53 VWAPGEGTAFKCLLSARLCAALLSNISDCDETFNWFPHYLIIYEGEGTWEYSPAIR 112

```
Db 15 VW-----TSLILAVRLTASFRVDDDEVYNYNEPLHYLYGLOQWYSREYAIR 68
Qy 113 SYAYLLHAPPAHARILQTNKILVYFLRCLLAFVSCICELFYKAVCKFGLHVRSM 172
Db 69 SWFYALHAVP-GFLARGGLSLRHVYFIRGLVACFSAPCETNLLILAVARNFNAVALH 127
Qy 173 MLAEVLSTGFCSSAFPLSSPCWYTTLTAMTGWY---MDKTSIAVLG-VAAGALLGWP 228
Db 128 LTVSVFNSGWSASTSFLSSPANNVUTLALSAQSPSTKTKVAVSVFTTIGAVIGWP 187
Qy 229 FSAALGLP-LAFDLVLMKRWKSP-HWSLWALILFLVP--VVVIDSYYYGKLVIAPLNI 284
Db 188 FSAALSIPFILLELDVLDKGRFHLPCRWKFAIFVALLITGIC-TVDSLFYHRIQFVAMNI 247
Qy 285 VLYNVFPP--HGPOLYTERKYFYLINGFLNFVAFALALLVPLTSLMEYLLQRFHQVN 342
Db 248 VKYNVLAKDGRGPDYGTGEPWYFYFANLSIQHNIVLWFAWACGLVLLAFLP----- 298
Qy 343 LGHPYWLTL-----APMYIWFITFIQPHKEERFLFPYYPPLIC:CGAVALSALQK 392
Db 299 ---TNWINDSFLDSSVISPPYIWLFIIFIQPHKEERFWYPIYVLCIAAIGLDMSLK 355
Qy 393 CYHFVQRYRLEHVTYS-----NWLALGTVPFLGCLLSFSRSVALFRGYHGLDLYPEFYR 448
Db 356 ---LMIQILSINETVRSKPVRVFLVCVVAIGCUSIARILAI-QYNAPMLIYPAISF 411
Qy 449 IATDPTHTVPEGRFNVNVCWKERYRFPSSFLPLDNWQLOFIPSEFRGOLPKPFAEGPL- 507
Db 412 LETDNVTT-----NVCVKWRYRPSFTFLPDNSRLKFKVSEFDGILPGEFVESNT 464
Qy 508 -----ATRVPTDNDQNLBPSYIDISKCHVLDL--DTMRETRPREKYSNKEEWS 560
Db 465 WNRREGYYQIPEHNEFNEPNEPTRYTSLSCDFLEDFDHSKATVNEPTYSKS-DGWNP 523
Qy 561 LAYRPFILDASSKLLARAFYVPLSDQYVYVNYTILKPKK 601
Db 524 VMVYPTIDTKTPMGRAFAPVFTPEKRWGY-BLVNKKPVK 563

RESULT 10
JS-09-922-225a-12
; Sequence 12, Application US/09922225A
; Publication No. US20030104385A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; TITLE OF INVENTION: Nucleic Acids and Encoded Polypeptides
; FILE REFERENCE: P-EA 4672
; CURRENT APPLICATION NUMBER: US/09/922.225A
; CURRENT FILING DATE: 2003-01-14
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 570
; TYPE: PRT
; ORGANISM: A. thaliana
US-09-922-225a-12

Query Match 24.8%; Score 812; DB 11; Length 570;
Best Local Similarity 32.6%; Pred. No. 8.1e-70;
Matches 197; Conservative 101; Mismatches 250; Indels 56; Gaps 13;

Qy 7 RQRLKGSAGSGDCTAPADKRLRELLGREGAGABHRTLSGNKAGQVWAPEGSTAFKCLL 66
Db 7 RQRRPLISDSSSSSTKSYKTDKPRSGNGDAB-----DGLRW----FLPFIALC 54
Qy 67 SARLCAALLSNISDCDEFNTEPHYLIIYEGQOTWEYSAPAIKRYAYVLLH---AMP 123
Db 55 YLYRYSATSNIIHDCDEFNTEPHYLIIYKSGFTWEYSNFAIRSYLYLTHELAGRP 114
Qy 124 AAFHARILQTNKILVYFLRCLLAFVSCICELFYKAVCKFGLHVRSMMLAFVLSTGM 183
Db 115 ASMW---FGDDKRVFYAVRULGLVSAVSDTLVVALSRKYGRKIATYAVAMCLTSGC 171
```

```
Qy 184 FCSSAFPLSSPCWYTTLTAMTGWMDKTSIAVLGVAAGAILGWPFSALGLPIADFLIV 243
Db 172 FFASTFLSSFSMAISLSSGLLFEKYAMAVSVVGVILGWFFSILAFPLPVVIVSLV 231
Qy 244 MKRWKSFHWSLMAILFLVPVWVIDSYYYGKLVIAPLNIVLVNVFTPHGPDLYGTBPW 303
Db 232 --KRPQAFIAGAVTIFLLGVSLLDVYYYKRWTSVNLNLLIYNVLGGESHLYGTEGA 289
Qy 304 FYLINGFLNFVAFALALLVPLTSLMEYLLQRFHQVNGLHPYWLTLAPMYIWFIIFFI 363
Db 290 LFYIRGFGNFNFCFTLAFVIAIYFVIRKRYDR-----ALLVVSVMYIWLAFMSL 341
Qy 364 QPHKEERFLFPVVPPLIC:CGAVALSALQKCYHFVQRYRLEHYTTSNWLALGTVFLFGL 423
Db 342 QPHKEERFLFPVVPPLIC:CGAVALSALQKCYHFVQRYRLEHYTTSNWLALGTVFLFGL 423
Qy 424 L-----SFSRSVALFRGYHGLDLYPEFYRIADPTHTVPEGRFNVNVCWKERYRPP 476
Db 394 ILGCILCASHSRFTFALINGYSAPLEVYKLE-----HDDAGFGSVLCVGSSEWHRYP 445
Qy 477 SSFLLPDN-WQLQFIPSEFRGOLPKPFAEGPLATRIPTDMDNDQNLBPSRYI-DISKCH 534
Db 446 SSFVPHYSEVANIDDGFRGLLPFFFNLTGTSASPPYFNKNQOASEEQYLNKNIETCT 505
Qy 535 YLVLDPTMRETPREPKYSNKEEWSLAYRPFILDASSKLLARAFYVPLSDQYVYVNY 594
Db 506 FLIELQLSR--PYQYR-GSDLTWSEIAVLPLYDRELSPAKYRSFFIPHMWQEKNVFGKY 562
Qy 595 TILK 598
Db 563 VALR 566

RESULT 11
US-10-291-265-240
; Sequence 240, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291.265
; CURRENT FILING DATE: 2000-01-25
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 240
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-240

Query Match 6.4%; Score 209.5; DB 12; Length 554;
Best Local Similarity 19.9%; Pred. No. 3e-11;
Matches 109; Conservative 85; Mismatches 196; Indels 159; Gaps 26;

Qy 87 YW---EPTHYLIYEGQOTWEYSAPAIKRYAYVLLHAPPAHARILQTNKILVYFLR 143
Db 87 YWQSLVSHWVFNFGYGLTWETE--RLASYIYPLIFASIIYKILHLHGKDSVQLLIWIPR 144
Qy 144 CLAFVSCICELFYKAVCKFGLHVRSMMLAFVLSTGMFSSSAFLPSSFCWYT---- 199
Db 145 LAQALLSAVDRLYSLMKQLEQNEVARWY-----FFCQLCSFWYWCCTRLTNT 195
```

Y 200 -----TLIANTGWYND-----KTSIAVLGVAAGAILGWPFSAALGLPIAFDOLLVWKHRWKS 250
 b 196 METVLITIALFYPLEGSKSMNSVKYSSSLVALAIIIRPTAVILWTLPLFRHFCQEPF--- 252
 Y 251 PFHMSMALILFLVVP-----WIDSYYGKLVIAPLNVLNVTTPHGPDLGYTEP 302
 b 253 -----KDLILHFLPVGVFTVLSLMIDRIFFGQWTLVQFNFLKFNVLQWNG-TFYGSH 307
 Y 303 WYFYLINGF---LNFNVAAPALALLVPLTSLMEYLLQRFHVNQNGHYPWTLAPMVIW-F 358
 b 308 WHWYFSQFPVILGTHLPFFI-----HGYLAPKRYRILLVTLVTL 349
 Y 359 IIFFIQPHKEERFLPVVFLICLGCAGVALSALQKCYHVFQRYLEHVTYSNWL--ALG 416
 b 350 LVISMLSHKEFRIFVPLP-FCM-----VFCYSLTH--LKWKPALS 390
 Y 417 TVPLFGLLSFSRSVALFRG---YHGPDLYPEFVRIATDPIHTHVPGRPVNV-----CV 468
 b 391 FLPLSNLF-----LALYGLVHQRGTLDVMSHIQKVC-----YNNPNKSSASIPIMPFCH 440
 Y 469 GKEMWFPSSFLPDNMQLOPTEPFRGOLPKPFAEGPLATRIVPTMDNDQNLERSYI 528
 b 441 STPY-----SHVHCLPMRFLQ-----CPDILT----- 464
 Y 529 DISKCHYLDLMTRETREPKYSSNKEEWISLAYRPFELASRSKULRAPYVFFLSQY 588
 b 465 --CKSHYLDADV-----FYLPLNLW---HREFHDDASLPTHLITPSI--LBEI 508
 Y 589 TVYV---NY 594
 b 509 SAFPSSNY 517

RESULT 12

S-10-080-170-547
 Sequence 547, Application US/10080170
 Publication No. US20030129601A1
 GENERAL INFORMATION:
 APPLICANT: COLE, S.T.
 TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 TREATMENT OF MYCOBACTERIOSES
 FILE REFERENCE: 03495.0218
 CURRENT APPLICATION NUMBER: US/10/080,170
 CURRENT FILING DATE: 2002-06-10
 PRIOR APPLICATION NUMBER: 60/270,123
 PRIOR FILING DATE: 2001-02-22
 NUMBER OF SEQ ID NOS: 652
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 547
 LENGTH: 433
 TYPE: PRT
 ORGANISM: Mycobacterium tuberculosis
 IS-10-080-170-547

Query Match 4.0%; Score 130; DB 16; Length 433;
 Best Local Similarity 21.5%; Pred. No. 0.0011;
 Matches 103; Conservative 62; Mismatches 169; Indels 144; Gaps 24;
 Y 18 GDTAPADKURELLGSRAGGAHRTLSGNKAGQWAPBEGSTAFKLLSARLCHALLS- 76
 b 3 GALVTRADSIRTGLGASLLAGPRPT-----GAPSTAT---ILRSALWPAANLSV 49
 Y 77 -----NTSD-----CDTFNWEPTHYLIYGEFQWTEYSP 107
 b 50 LHSIVLTNGNITDDPKPVYRAVLNFRGWDIYNEHFDYVP-HYL-----YPP 98
 Y 108 AYAIRSVAYLLHLAWPAAPFARILQNKILVFLFCLLAFCVSCICELYFYKAVCKEGL 167
 b 99 GGTLMAFFGFLPAPSRYLFIISNTAAILVAYL-----LLRMPNF 140
 Y 168 HVSRMMLAEVLSTGMFCSSS-----AFLPSSPCMYTTLIAMTGWMDKTSIAVLGVAAG 222

Db 141 TLTSAAPALILA--MPATETVNTVFTVNINGCILLLEVLFLRWLLD-----GR 188
 QY 223 AILGWPFSAALGL-----PIAFDLL---VWKHRWKSFFHWSMALILFLVPVVDSDYY 274
 Db 189 ASRQWCGGLAIGTLVLKPLGPELLPLINRQR-----ALVAAV--VVPVV----- 235
 QY 275 GKLVIAPLNVLNVTTPHGPDLGYTEPWFYFYLINGFLNFAFA-----LALLVLP 326
 Db 236 --NVAALFLVSDPMSFFRTLPYLGTED---YFNSILGNVYFGLPTWLLLEFLRLFTA 291
 QY 327 LTSMEYLLQRFHVNQNGH-PYWTIAP-----MYIWIIFFIQPHKEERFLPVVFLICL 381
 Db 292 ITFGALMLLYRY--RTGDPLEWFTTSSGVLILWSLWMSLAQY-YSMMLPFLMTVVL 348
 QY 382 -----CGAVALSALQKCYHVFQRY--RLBHYVTYSNWLALGTVFLFGLISF 426
 Db 349 PNVIRNWPALVGYGVFWTLDRMLLFWNMGRALEYLKITYGM-SLLIVITVTVLYF 405

RESULT 13

US-10-032-585-7898
 Sequence 7898, Application US/10032585
 Publication No. US20030180953A1
 GENERAL INFORMATION:
 APPLICANT: Terry, Roemer D.
 APPLICANT: Bo, Jiang
 APPLICANT: Charles, Boone
 APPLICANT: Howard, Bussey
 TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
 FILE REFERENCE: 10182-005-999
 CURRENT APPLICATION NUMBER: US/10/032,585
 CURRENT FILING DATE: 2001-12-20
 NUMBER OF SEQ ID NOS: 8000
 SOFTWARE: Patent In version 3.1
 SEQ ID NO 7898
 LENGTH: 584
 TYPE: PRT
 ORGANISM: Candida albicans
 US-10-032-585-7898

Query Match 3.8%; Score 126; DB 12; Length 584;
 Best Local Similarity 21.2%; Pred. No. 0.0042;
 Matches 97; Conservative 69; Mismatches 149; Indels 142; Gaps 25;
 QY 198 YTTLIAMTG---WYMDKT-----STAVLGVAAG---AILGWPFSAALGLP 236
 Db 139 YTTLLSQFHLIYSSRTLDPNFIALPLNFSLSKLIQGLSLTWTWLAFTGIVFLEVLG- 197
 QY 237 IAFDILLVWKHRWKSFFHWSMALILFLVPVV-----IDSYYGKLVIAPLNVLN 288
 Db 198 --FGLIIAIVSSLSGFGQSNIFGNIYILAMGTLGGT*SFICIDSYFVWGRPLIPEIDSIFN 255
 QY 289 VFTPHGPDLYGTEPW--YF--YLINGFLNFAVALALLVPLTSLMEYLLQ--RF-HV 340
 Db 256 IVQKSTE-WGTEPMDTYFKKYL-----FQLEPPVILMLAIPGLINDPANDGTGFGDK 308
 QY 341 QNLGHYPWTLAPYINFIIF-----IQPHKEERFLPVVFLICL-----CGAVAL 387
 Db 309 KSVPHARYSLNLFISILLFIAAMSFPQHKWRFIVVTIPITF*TLQAANGVTN*CKWGL 368
 QY 388 SALQKCYHVFQRYLEHVTYSNWLALGTVFLGILLSFSRSVALFRYHGPDLGYPEFY 447
 Db 369 SVLNKVLFIIG-----ANVTISLLS-----LHMYISSFN-YF--- 402
 QY 448 RIATDPTIHTPEGRPVNVCVKWEYRFPSSFLPDNMQLOF-----IFSEFRGQLPKP 501
 Db 403 -----GGDALQFTNNVIL-ENYKNETVSVHMDVPACMTG----- 435
 QY 502 PABGPLATRIVPTMDNDQNLERSYIDISKCHYLDLDTWRETR-EPKYSSNKEEWIS 560
 Db 436 -----ITRFGELDKGFASYDKSEQDFDITNYDIIII---THNEVENWELLHSSRVFGIS 486
 QY 561 LAYRPFELASRSKULRAPYVFPFLSDQYTVVYNTIL 597

db 487 L-RMFIQIFMAQR-----KDRSHIP-NITKL 510

RESULT 14

US-10-080-170-164

Sequence 164, Application US/10080170

Publication No. US20030129601A1

GENERAL INFORMATION:

APPLICANT: COLE, S.T.

TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR

TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR

TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES

FILE REFERENCE: 03495.0218

CURRENT FILING DATE: 2002-06-10

PRIOR FILING DATE: 2001-02-22

NUMBER OF SEQ ID NOS: 652

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 164

LENGTH: 440

TYPE: PR

ORGANISM: Mycobacterium leprae

US-10-080-170-164

Query Match 3.6%; Score 117; DB 16; Length 440;

Best Local Similarity 20.6%; Pred. No. 0.021;

Matches 80; Conservative 59; Mismatches 146; Indels 104; Gaps 20;

82 DETFNWEPHYLLYIGEGFQWTSYSPAYAIRSYAYLLHAWPAAPHARILQTNKILVEYF 141

84 NEHPDYVDP-HYL-----YPPGGLTLLMAPFGYLPAPSRVLFILINTGALLIAY 132

142 LRCILAFVSCICELYFYKAVCKFGLHVSRRMLAFVLSTGMFC-----SSSAFLPSSFC 196

133 L-----ILRLPKYTLSSVAAPTLLA--MFCETVTSTLVTFTNGC 172

197 MYTILAMTGYMDKTSIAVLGVAAGAILGWPFSAALGL-----PLAFDLL---VMKRW 248

173 IMLEVLFLMWLINGSFPTVSOQ-----WAGGAIGLTLVKPLGLPLCLPLINRQW 226

249 KSFFHWSMALI-LFLVPVVVIDSYVYKGLVIAPLNVLNVFTPHGPDLYGTEPW---- 303

227 Q-----ALVPAIALPVVI-----NLAALPLVSHPMDFTRVPVILGTRDYFNSS 271

304 -----YFYLINGFLNVAFAFALLVPLTSLMEYLLQRFVONLGHYPYWLTP----- 353

272 IEGNGVYFGLPTWLIVF---LRLFTVLAICSL--WLLARYY-RTDRPLFWFTCTGVLL 325

354 MYIWFIFFIQPHKEERFLFPVYPLICL-----CGAVALSALOKCYHFVFORQY- 401

326 LWSLVLPLAQGY-YSNMLEFFFLMTVVLPNSLNRNWPAMLIYGFILDRWLLFNWMEYG 384

402 -RLEHYTVTSNW-----LALGTVFLFGLL 424

385 RALEYKITYGWSLLIIVSVTLCFRYL 412

RESULT 15

US-09-862-767A-9

Sequence 9, Application US/09862767A

Patent No. US20020034786A1

GENERAL INFORMATION:

APPLICANT: Pan, Yang

TITLE OF INVENTION: TANGO-72 AND TANGO-77 NUCLEIC ACID MOLECULES AND POLYPEPTIDES

FILE REFERENCE: MSIO1997-0081RCPICN1(M)

CURRENT APPLICATION NUMBER: US/09/862,767A

PRIOR FILING DATE: 2001-05-21

PRIOR FILING DATE: 1998-08-03

PRIOR FILING DATE: 1997-08-04

NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 458
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-862-767A-9

Query Match 3.4%; Score 111.5; DB 9; Length 458;
Best Local Similarity 21.6%; Pred. No. 0.075;
Matches 78; Conservative 50; Mismatches 130; Indels 103; Gaps 20;

QY 105 YSPAVAIRSYAYVLLHAWPAAPHA-----RILOTN--KILVYFLRCLLAFVSICE 154

DB 116 FNPCEIDMGYNILRVLIWFISILAITGNIIVLVITTSQYKLTVPFLMCLNLAFLAD-LC- 173

QY 155 LYFYKAVCKKFGHVSRRMLAFVLV-LSTGMFCSSAF--LPSSFCMYT-TLIAMTGWY- 208

DB 174 IGIYLLLIASVDIHTKSYHNVYDAIDWQIGAGDRAAGFTVFASLSVYTLTATLERWHT 233

QY 209 -----MDKTSIAVLGVAAGAILGWPFSAALGLPLAFDILLVMKRWKSFHWSL----- 256

DB 234 ITHAMQLD-CKVQLRHAASVMVMGWIFAFAAALPFIFGI-----SSYMKVSICLPMDI 285

QY 257 -----WAILLP--LVPVVVIDSYVYKGLVIAPLNVLVN-----VETPH 293

DB 286 DSPLSOLYMSLLVLNVLAFVVICGCIHIYLVNPNIVSSSDTRIAKRMAMLIFF-- 343

QY 294 GPDLYGTEPWYFYLINGFLNVAFAFALLVPL--TSLMEYLLQRFH-VONLGHYPYWL 350

DB 344 --DFLCWAPISFFAIS-----ASLKVPLITVSKAKILLVLPHPINSCANPFLYA 390

QY 351 LAPMYIWFIFFIQPHKEERFLFPVYPLICGVALSALOKCYHFVFORVLESHYTVTS 410

DB 391 I-----FTKNFRDRDFFI-----LLSKCG-----CYEMOQAIYRTTSTSTVH 426

QY 411 N 411

DB 427 N 427

Search completed: January 13, 2004, 07:09:54
Job time : 56 secs